Aalto University
School of Science
Degree Programme in Engineering Physics and Mathematics

Anu Harjula

Time Series Modelling of Exchange Rates

Master's thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Technology in the Degree Programme in Engineering Physics and Mathematics.

Espoo, 3rd of January 2013

Supervisors: Professor Esko Valkeila & Professor Olavi Nevanlinna
Instructors: Ilkka Mellin, Lic.Soc.Sc. & Helena Aro, M.Sc. (Tech.)

Aalto University
School of Science

ABSTRACT OF THE MASTER'S THESIS

Author: Anu Harjula

Title: Time Series Modelling of Exchange Rates

Title in Finnish: Valuuttakurssien aikasarjamallintaminen

Degree Programme: Degree Programme in Engineering Physics and Mathematics

Major subject: mathematics | Minor subject: economics

Chair (code): Mat-1

Supervisors: Professor Esko Valkeila, Professor Olavi Nevanlinna | Instructors: Ilkka Mellin, Lic.Soc.Sc., Helena Aro, M.Sc. (Tech.)

Modelling volatility is important since it relates closely to risk, which is a key topic in economics. If the volatility of a time series varies over time it is said to have heteroskedastic effects. The models of heteroskedasticity have been introduced quite recently (Engle 1982), and they have been shown to be very effective even with few parameters. The aim of this thesis is to find sufficiently good models for three exchange rate datasets, and we are especially interested in modelling the heteroskedasticity which can be seen as volatility clustering in the data.

We explore the time series modelling of three exchange rate series: the United States dollar, the Swedish krona and the pound sterling each compared to the euro. First we use autoregressive moving average (ARMA) models to remove the serial correlations in the time series. Then we apply ARCH (autoregressive conditional heteroskedasticity) and GARCH (generalized autoregressive conditional heteroskedasticity) models to the residuals of the chosen ARMA models to dispose of the volatility clustering.

We start by discussing the importance of modelling volatility in economics. Then we proceed to the data analysis. Several ARMA models are estimated for the three datasets to remove the serial correlations. We compare the ARMA models by means of two information criteria (Akaike information criterion and Bayesian information criterion) to find the best fit. Next we apply simple ARCH and GARCH models to the residuals of the chosen ARMA models. The models of heteroskedasticity are evaluated using graphical residual analysis.

We find ARMA models that are good fits to each dataset. i.e., the models remove all the significant serial correlations. In addition, the GARCH(1,1) model turns out to be an accurate model for the heteroskedasticity in all three series and it manages to even out the volatility clustering.

Date: 3rd of January 2013 | Language: English | Number of pages: iv + 55

Keywords: ARMA, ARCH, GARCH, heteroskedasticity, volatility, volatility clustering, time series modelling

Aalto-yliopisto Perustieteiden korkeakoulu	DIPLOMITYÖN TIIVISTELMÄ

Tekijä: Anu Harjula

Työn nimi: Valuuttakurssien aikasarjamallintaminen

Title in English: Time Series Modelling of Exchange Rates

Tutkinto-ohjelma: Teknillisen fysiikan ja matematiikan tutkinto-ohjelma

Pääaine: matematiikka	Sivuaine: kansantaloustiede

Opetusyksikön (ent. professuuri) koodi: Mat-1

Työn valvojat: Professori Esko Valkeila, Professori Olavi Nevanlinna	Työn ohjaajat: Ilkka Mellin, VTL, Helena Aro, DI

Volatiliteetin mallintaminen on tärkeää, sillä volatiliteetin ja riskin välillä on suora yhteys. Riski taas on yksi kiinnostavimmista aiheista taloustieteessä. Jos aikasarjan volatiliteetti vaihtelee, sanotaan siinä olevan heteroskedastisuutta. Mallit heteroskedastisuudelle ovat melko tuoreita (Engle 1982), ja ne ovat todistetusti hyvin tehokkaita jo muutamilla parametreilla. Tämän diplomityön tavoitteena on löytää tyydyttävät aikasarjamallit kolmen valuuttakurssin mallintamiseen ja keskitymme erityisesti mallintamaan heteroskedastisuutta, joka näkyy aikasarjoissa volatiliteetin kasaantumisena.

Tarkasteltavat valuuttakurssit ovat Yhdysvaltain dollarin, Ruotsin kruunun ja Englannin punnan kurssit suhteessa euroon. Aluksi sovellamme aikasarjoihin ARMA-malleja (autoregressive moving average models), joilla mallinnetaan sarjakorrelaatioita valuuttakursseissa. Tämän jälkeen käytämme ARCH- ja GARCH-malleja (autoregressive conditional heteroskedasticity models, generalized autoregressive conditional heteroskedasticity models) ARMA-mallien residuaaleille volatiliteetin kasaantumisen poistamiseksi.

Aloitamme keskustelulla volatiliteetin mallintamisen tärkeydestä taloustieteessä, jonka jälkeen siirrymme aikasarjojen analysointiin. Sovellamme useita ARMA-malleja jokaiseen valuuttakurssiaikasarjaan sarjakorrelaation poistamiseksi. ARMA-mallien vertailussa käytetään kahta informaatiokriteeriä (Akaike information criterion, Bayesian information criterion). Seuraavaksi sovellamme yksinkertaisia ARCH- ja GARCH-malleja valittujen ARMA-mallien residuaaleihin. Malleja heteroskedastisuudelle arvioidaan graafisen residuaalianalyysin avulla.

Onnistumme löytämään jokaiselle kolmesta valuuttakurssista sopivan ARMA-mallin, joka poistaa kaikki merkittävät sarjakorrelaatiot. Lisäksi, GARCH(1,1)-malli osoittautuu tarkaksi malliksi aikasarjojen heteroskedastisuudelle, ja se tasoittaa kaiken volatiliteetin kasaantumisen.

Päivämäärä: 3.1.2013	Kieli: englanti	Sivumäärä: iv + 55

Avainsanat: ARMA, ARCH, GARCH, heteroskedastisuus, volatiliteetti, volatiliteetin kasaantuminen, aikasarjamallintaminen

Acknowledgements

Many people have participated in finishing this thesis, and I would like to take this opportunity to show my appreciation. Firstly, I would like to thank my original supervisor, late professor Esko Valkeila for all his guidance in the beginning. He helped with specifying the subject and getting the work started. I would also like to thank professor Olavi Nevanlinna for taking over as my supervisor and guiding the work to its final version.

In addition, I would like to thank my instructors Ilkka Mellin and Helena Aro for dedicating so much of their time and effort to helping me with this thesis. Their comments and guidance have been an invaluable resource throughout the working process.

I would also like to thank my family and friends who were an irreplaceable source of encouragement and support. Special thanks to Leif who was always there for me.

Anu Harjula
Espoo, 3.1.2013

Contents

1 Introduction

Risk is one of the most important aspects of any financial market. It is also the most difficult part to predict since it involves human behavior and thus sometimes irrational decisions. Modelling risk has long been a key field of study in economics and the recent financial crisis showed that there is still a lot to learn in that area. Risk is closely related to volatility, by which we mean fluctuations around the trend line of an economic time series. Hence, understanding the dependencies in volatility is of particular interest.

The changes in volatility over time are called the heteroskedasticity effect and similarly a series with no changes in volatility is called homoskedastic. Heteroskedasticity often appears as volatility clustering when higher variances occur together. Similarly, times of low volatility, and thus low risk, are followed by similar times. If we had an accurate way of modelling heteroskedasticity it would allow us to predict time periods of higher risk and thus be prepared for larger losses. This is something every investor is trying to accomplish in order to be able to decide when to buy and sell their assets.

The aim of this thesis is to model the dependencies in exchange rate time series using models of heteroskedasticity. We observe the similarities and differences in the time series and their dependencies on each other. A secondary objective is to demonstrate the behavior of an economic time series in general.

Possible ways to model heteroskedasticity have been studied only during the past few decades and the breakthrough in this field was the autoregressive conditional heteroskedasticity (ARCH) model introduced in the seminal work of Engle [Engle, 1982]. He was the first to formally recognize the dependence in the variances of an economic time series. Serial correlations had been modelled before using, e.g., autoregressive moving average (ARMA) models but Engle extended this idea to the variances. The work of Engle was continued by Bollerslev who took the idea of the ARCH model further and presented the generalized conditional heteroskedasticity (GARCH) model [Bollerslev, 1986]. After this many variations of ARCH and GARCH have been developed, but still these fairly simple models continue to successfully capture the heteroskedasticity in many economic time series.

In Section 2 of this thesis we discuss some aspects of volatility in economics and explain more closely why modelling volatility is of particular interest. Then, in Section 3 we take three exchange rates, namely the United States dollar, the Swedish krona and the pound sterling, each compared to the euro. After some preprocessing we model the serial correlations in each dataset with autoregressive moving average (ARMA) models and aim to find a model that is a good fit to the data. We use the Akaike information criterion and the Bayesian information criterion to determine the suitable orders for the ARMA models. Then we examine the graphical residual analysis, containing also the correlogram and the partial correlogram of the residuals, for the best ARMA models to see whether the models are good fits to the datasets and succeed in removing the significant serial correlations.

Next, in Subsection 3.4 we proceed to study the heteroskedasticity in the three time series and look for evidence for that models of heteroskedasticity should be applied. Finally, we apply some simple autoregressive conditional heteroskedasticity (ARCH) models and generalized autoregressive conditional heteroskedasticity (GARCH) models to the residuals of the chosen ARMA models for each dataset. To study the goodness of fit of these models we again look at the residual analyses.

2 Volatility in economic time series

Especially in economic time series volatility is often a point of interest. The volatility of an exchange rate refers to the non-constant fluctuation of the exchange rate around a constant mean or a trend line. This kind of variations are due to revelations of new information to the financial market. Volatility is usually measured as the conditional standard deviation of the underlying asset return, here the exchange rate [Tsay, 2010]. It is considered to be the most important measure of risk in the financial market and thus it is an interesting research subject [Bekaert and Wu, 2000].

Volatility corresponds with risk quite straightforwardly. If the price of the asset changes substantially the risk of loss gets higher for the owners of these assets. It is also worthwhile to monitor the changes in volatility. When the asset price varies only little, i.e., there is little volatility, then there is less uncertainty and the risk is considered smaller. But when volatility suddenly increases, the risk of loss increases also. This often causes the volatility to stay high for a longer period of time due to the uncertainty, which can be seen as significant drops and increases in the price following each other closely.

Volatility is a suitable and efficient measure of risk for two reasons. First of all, it is very simple to measure and compare the volatilities of assets. And secondly, volatility gives a clear way to study the behavior of the asset price from the perspective of earning and losing money. If the volatility is high, the investors are more likely to have big winnings but also big losses. When the volatility is lower, both the winnings and the losses stay moderate. This is exactly the issue investors face when deciding whether to take risks.

Traditionally in economics the volatility has been considered to be a constant and hence something that cannot help in forecasting. However, during the last few decades it has been discovered that many economic time series have non-constant variances [Hendry and Juselius, 2000]. This suggests that there is time dependence and predictability in the variances. It is quite intuitive that at uncertain times (as during the financial crisis in 2008) the exchange rate varies more than at economically more stable times. Major events and the overall state of the economy affect the exchange rate differently at different times and thus cause the volatility to vary.

The theory of dependent volatility led to the concept of volatility clustering, i.e., the higher variances tend to occur together. If the exchange rate has remained steady it most probably will stay that way. And usually large drops are followed by a relatively large increase, and vice versa. This kind of behavior can be modelled with allowing the variance of a data point to depend somehow on the past variances (or prediction errors). Heteroskedastic models presented later in this work (Appendix A.7) have this property.

As many studies [Ball and Mankiw, 1994, Cover, 1992] show, the financial market does not react similarly to positive and negative shocks. This is called the leverage effect and is due to the fact that the trust of the investors is much easier to lose than to gain. However, the most common heteroskedastic models (e.g., ARCH and GARCH in Appendices A.7.1 and A.7.2) cannot take into account this difference, which leads to prediction errors. There does exist more complex models of heteroskedasticity which can model the asymmetricity of shocks (Appendix A.7.3), but they are not utilized in this work.

In addition to the issue of volatility clustering, economic time series are often heavy tailed, i.e., there is more probability mass at the tails than in the normal distribution [Loretan and Phillips, 1994]. This is due to the occasional unusually high or low values that appear when something unexpected happens in the market. These extreme data points make the distribution to have more mass at the tails. Due to this property of heavy tails, the residuals of economic time series models are often t-distributed instead of being normally distributed.

Volatility exists in every area of finance and economics, not only in the exchange rate market. Stock prices and government loans are just as vulnerable to the changes in the prices or the interest rates and economists spend a lot of time and effort on managing the risks caused by volatility.

The prices in the stock market are determined by supply and demand, and the price of today is always equal to the expected future value of the asset [Parkin et al., 2008]. If the agents performing in the market think that the value of the asset is going to rise they want to buy the asset, and if they think the value is going to drop they want to sell it. However, when most of the agents have the same opinion and everyone wants to buy the asset for example, this increases the current price of the asset. Due to this effect the price of today balances to equal the expected price in the future.

These predictions of the changes in stock price make the price of the asset to be very volatile especially on economically more uncertain times. The price can fluctuate significantly and often large drops are followed by another drop or an unusually large increase. This volatility clustering is a phenomenon often seen in the stock market but also in the exchange rate market.

The financial crisis in 2008 was a good example of how much the changes in volatility affect the market prices and the whole economy. The chain reaction began with the irresponsible granting of mortgages in the US, which eventually led to the real estate bubble bursting. This made the interest rates to drop which again increased the uncertainty, volatility and risk in both stock and exchange rate markets. The effects were so drastic since the agents working with finance are human. Everyone tries to protect their own investments and become more cautious when there is more uncertainty in the market. Thus, the changes in the asset price are more volatile than if the system was handled by totally rational agents. Also, the agents usually react more aggressively when there is risk for big losses than for big winnings. This effect is called market psychology and it needs to be taken into account when making economic analysis.

As discussed in this section, volatility is a very important part of the analysis of financial time series. Next we will take three real datasets and explain their behavior using the ideas of volatility introduced in this section. We also aim to find reasonably good time series models for the data.

3 Data analysis

In this section we take the three exchange rate time series and model them first with autoregressive moving average (ARMA) models. Then we apply models of heteroskedasticity to the residuals of the chosen ARMA models. Finally, we divide one of the datasets into two parts and model these parts separately to possibly obtain better results.

3.1 Presentation of the data

The datasets used in this study are the exchange rates of three currencies: the United States dollar, the Swedish krona and the pound sterling, compared to the euro. The data are downloaded from the website of the European Central Bank (http://www.ecb.int/stats/exchange/eurofxref).

The exchange rate data are for the period from January 4^{th} 1999, when the euro was first established, until July 3^{rd} 2012. The rate between the U.S. dollar (USD) and the euro is presented in Figure 1, the rate between the Swedish krona (SEK) and the euro in Figure 2 and the rate between the pound sterling (GBP) and the euro in Figure 3. Just by examining these three graphs we see that there are no clear outliers or missing data to be dealt with. The exchange rates are only calculated on weekdays, hence there are no data for weekends or for bank holidays. There are 5 trading days in a week and approximately 252 trading days in a year.

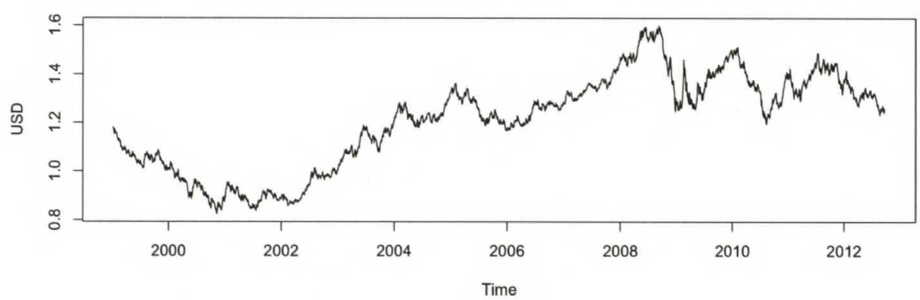

Figure 1: The exchange rate between the U.S. dollar and the euro from 1999 to 2012

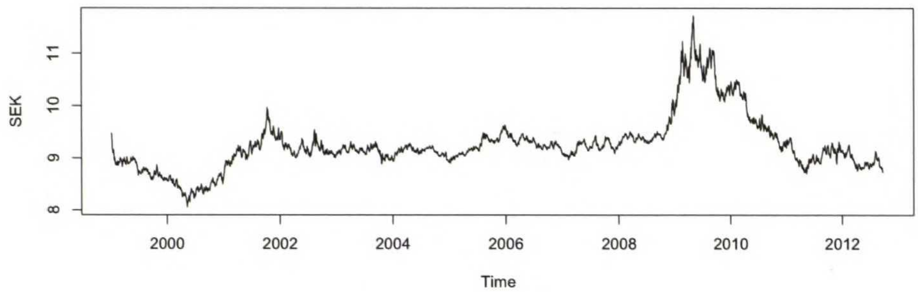

Figure 2: The exchange rate between the Swedish krona and the euro from 1999 to 2012

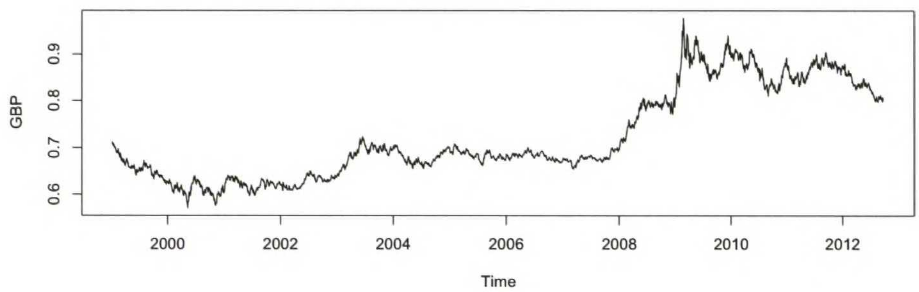

Figure 3: The exchange rate between the pound sterling and the euro from 1999 to 2012

All of the three exchange rate data (Figures 1, 2 and 3) are very volatile and no monotonic trends are visible. The USD data first declines until the beginning of 2002, reaching its minimum value of 0.8525 in October 2000. Then the rate starts to climb until it reaches its highest value of 1.5990 in July 2008. Shortly after this, there is a sudden drop due to the beginning of the financial crisis and since then, as the crisis affected Europe also, the rate has been fluctuating around the same value.

The SEK and GBP rates have evolved quite similarly over the observation period. There is a slight drop in the beginning of both time series when the euro devalued just after being introduced. The rates reach their minimum values in May 2000 (8.055 for SEK and 0.5711 for GBP). Next, there is quite a long period where the exchange rates of the euro against the krona and the pound stay fairly steady. However, when the financial crisis hit, these two rates increased significantly and started to fluctuate strongly. The economy of Great Britain was affected earlier as the GBP rate starts to rise in the beginning of 2008 and the SEK rate only in the beginning of 2009. The pound rate reaches its maximal value of 0.97855 in December 2008. The krona rate is at its highest of 11.7135 in March 2009.

In 2011 the SEK rate drops under its value before the crisis, but the GBP rate stays high until the end of the dataset and shows only minor signs of dropping during the year 2012. This suggests that the Swedish economy recovered from the crisis much faster than the euro zone as Great Britain continued to struggle.

One significant feature of all of the three datasets is the volatility clustering, i.e., the higher variances occur together. When there is a large peak it is often followed by a large drop (as happened during the financial crisis) and when the series has been fluctuating very little it most often stays that way. Another interesting point is that the values of the time series seem to follow the previous values rather closely. This suggests that there are some serial correlations in the data.

In Figure 4 are the original USD time series as well as its QQ-plot against the normal distribution, and autocorrelation (ACF) and partial autocorrelation (PACF) functions (Appendix A.2). In Figures 5 and 6 are the same graphs for the krona and the pound, respectively. From the QQ-plots we can see that none of the series is normally distributed. However, the USD and GBP datasets seem to have fairly symmetric distributions. The SEK data, on the other hand, is not symmetrically distributed. All three distributions have heavier tails than the normal distribution, which could suggest them following Student's t-distribution. In addition, the shapes of the ACF and PACF graphs of all three rates indicate that the time series are non-stationary. The dashed horizontal lines in the ACF and PACF graphs are at $\pm 2/\sqrt{n}$, where n is the number of observations. If the values are outside these boundaries, it suggests that there are serial correlations at that lag (see Appendix A.2).

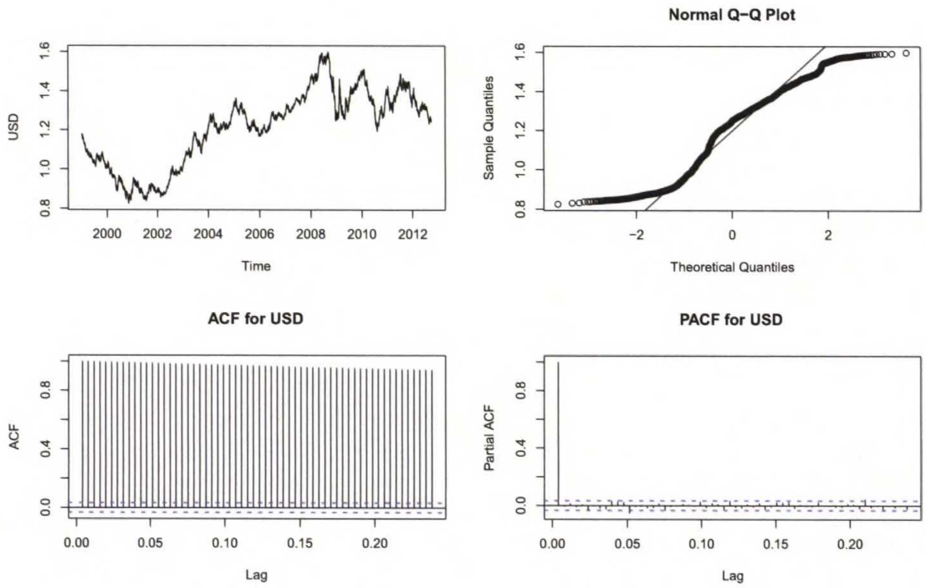

Figure 4: The U.S. dollar time series and its normal QQ-plot, correlogram and partial correlogram.

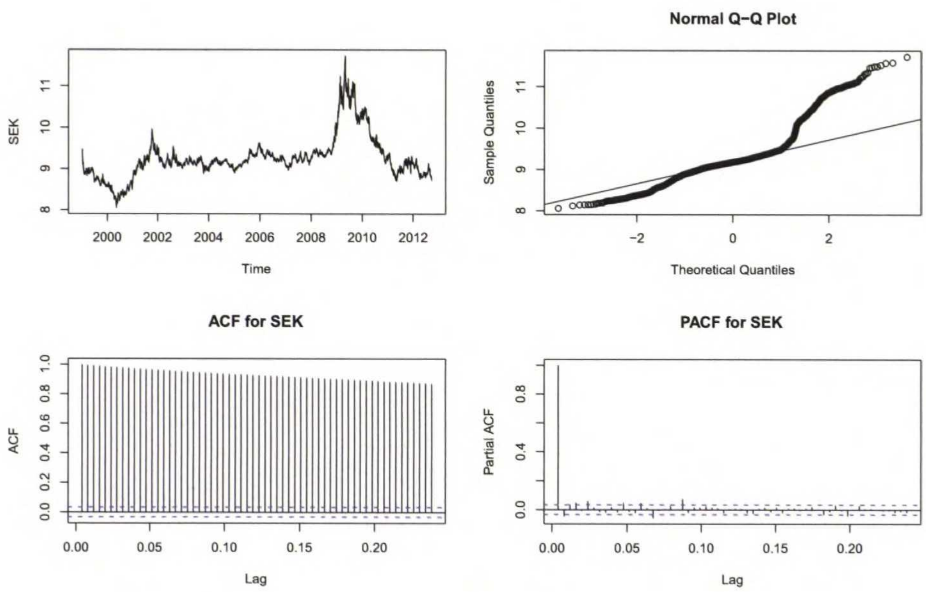

Figure 5: The Swedish krona time series and its normal QQ-plot, correlogram and partial correlogram.

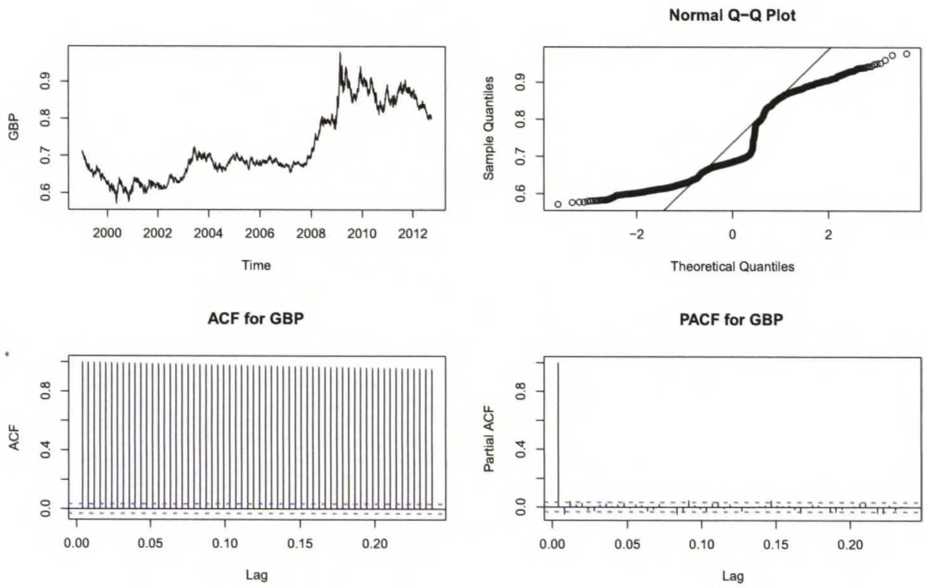

Figure 6: The pound sterling time series and its normal QQ-plot, correlogram and partial correlogram.

Based on the observations made in this section, we start modelling the data separately with autoregressive moving average models (see Appendix A.3) to deal with the serial correlations. Then we apply models of heteroskedasticity (see Appendix A.7) to the residuals of the chosen ARMA models to remove the volatility clustering. The ARMA models are applied before the heteroskedastic models since the serial correlation in the data would interfere in applying the heteroskedastic models. But first we do some preprocessing for the data.

3.2 Preprocessing of the data

Before actually starting we apply some general preprocessing methods in order to make the data more convenient for modelling. Especially we aim to make the time series more stationary, since the models used later in this study suit only stationary data. A time series is stationary if its mean and variance stay unchanged throughout the data and the correlation between points x_t and x_s depends only on the time difference $(s-t)$. The procedures tested include taking the natural logarithm of the data and the first difference of the data, i.e., $Y_t = X_t - X_{t-1}$, where X_t is the original series.

Also differencing the logarithmic series is tested.

Taking the natural logarithm of the series alone did not make any of the three datasets more stationary or otherwise simpler to model and we move on to differencing the data. Figure 7 shows the once differenced series, a normal QQ-plot and the correlogram and partial correlogram of the differenced USD data. The graph shows that the once differenced series is much more stationary than the original series, since its mean stays approximately the same. The variance fluctuates though, which is a sign of the need for heteroskedastic models (Appendix A.7). The shape of the QQ-plot suggests that the series is not normally distributed either, but that its distribution has heavier tails than the normal distribution. The distribution seems to be quite symmetric, which together with the heavy tails points to a Student's t-distribution. The ACF and PACF suggest that there are serial correlations at lags 9, 12, 29 and 52.

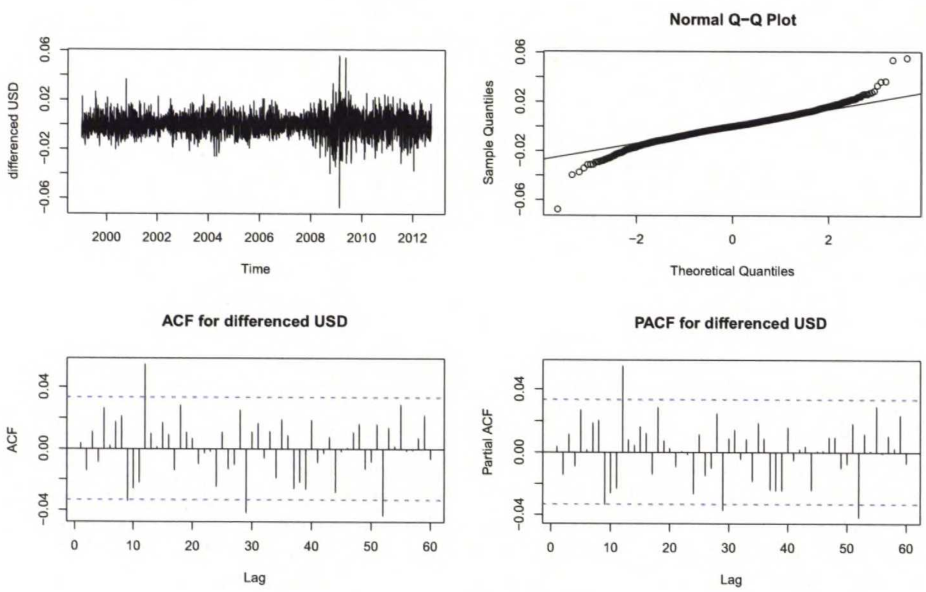

Figure 7: Differenced USD series, QQ-plot and correlogram and partial correlogram for the differenced series

We also tried taking the difference of the natural logarithm of the original USD series. At least with naked eye it did not seem to do anything to the time series compared to the once differenced USD series. Differenced logarithm had slightly less serial correlations when we looked at the ACF and PACF, although the only real difference was that the peak at lag 9 was not significant for the logarithmic difference of the USD series.

When examining the SEK and GBP datasets we found that the differences of the logarithmic series seemed to be more stationary than the differences of the original series. Thus, we choose the differenced logarithmic series for these two datasets. In Figures 8 and 9 are the differenced logarithmic series for the SEK and GBP data, respectively. The figures also include a QQ-plot and the ACF and PACF of the differenced logarithmic data. According to these figures neither the SEK nor the GBP data is normally distributed and there are significant serial correlations in the datasets.

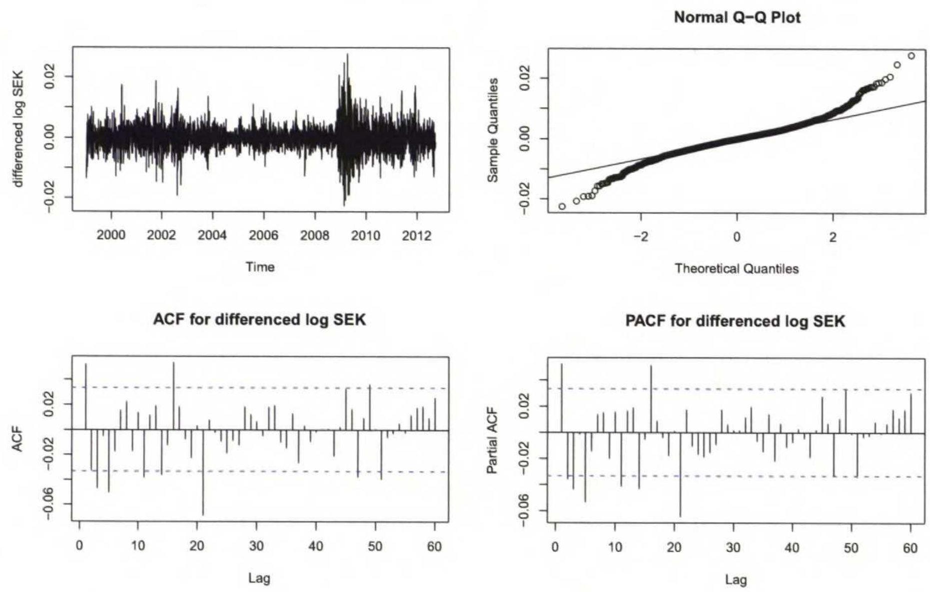

Figure 8: Difference of logarithmic SEK series, QQ-plot and correlogram and partial correlogram for the differenced series

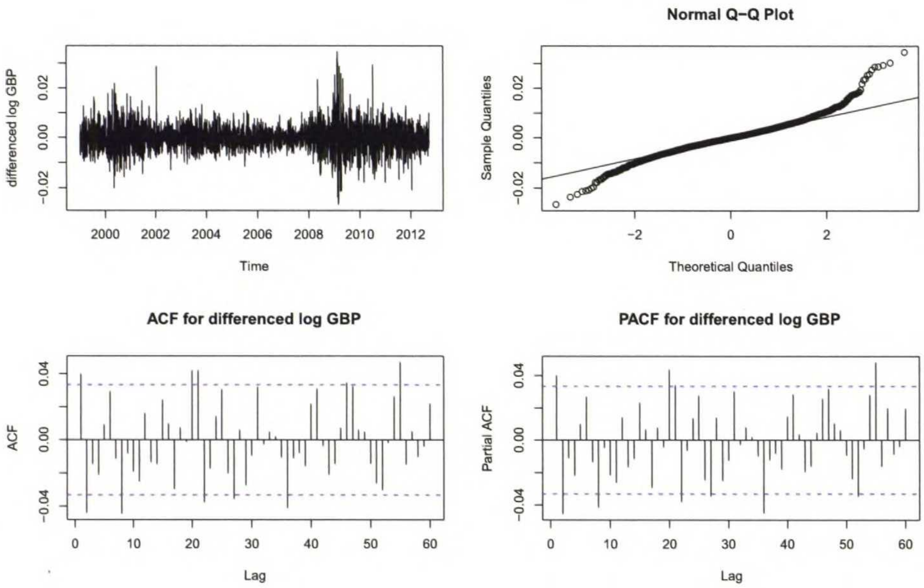

Figure 9: Difference of logarithmic GBP series, QQ-plot and correlogram and partial correlogram for the differenced series

When comparing the differenced USD series (Figure 7) with the same analysis of the original USD data (Figure 4) its is clear that the differenced USD series is more stationary and there is significantly less serial correlation. Thus, working with the once differenced USD series should be more sensible than fitting models to the original USD data. The same holds for the other two time series. In Figures 8 and 9 (differenced logarithmic SEK and GBP series) there are less serial correlations than in Figures 5 and 6 (original SEK and GBP series). We decide to continue working with the once differenced USD series and the differenced logarithmic SEK and GBP series, since they are more stationary than the original series.

3.3 Estimating ARMA models

In this section we estimate some autoregressive moving average models (Appendix A.3) for the three datasets to remove the serial correlations in them. The ACF and PACF of the differenced USD data presented in Figure 7 show that there are significant correlations at lags 9, 12, 29 and 52. The peak at lag 9 is barely significant but the peak at lag 12 is quite large. This could mean that an ARMA model, with less than 12 parameters cannot do well in modelling the data. The SEK and the GBP datasets also have significant serial correlations at several lags as can be seen in the ACF and PACF graphs of Figures 8 and 9. For the rest of this section we call the differenced USD series, the differenced logarithmic SEK series and the differenced logarithmic GBP series solely the USD, the SEK and the GBP series, respectively.

We start by modelling the three time series with autoregressive moving average models (ARMA(p, q)) where $p, q = 1, 2, \ldots, 20$ and $p + q \leq 21$. Having more parameters would make the model too complicated and to have too many insignificant parameters. Later in this section we estimate ARMA models with only the terms corresponding to the peaks in the ACF and PACF plots. To compare the ARMA models we calculate their Akaike information criteria (AIC) (cf. Appendix A.5). The lower is the AIC the better fit the model is to the data. Since the AIC includes also a penalty term the resulting model will not become too complicated, i.e., have too many parameters. The AIC values of the estimated ARMA models are shown in Figures 10, 11 and 12.

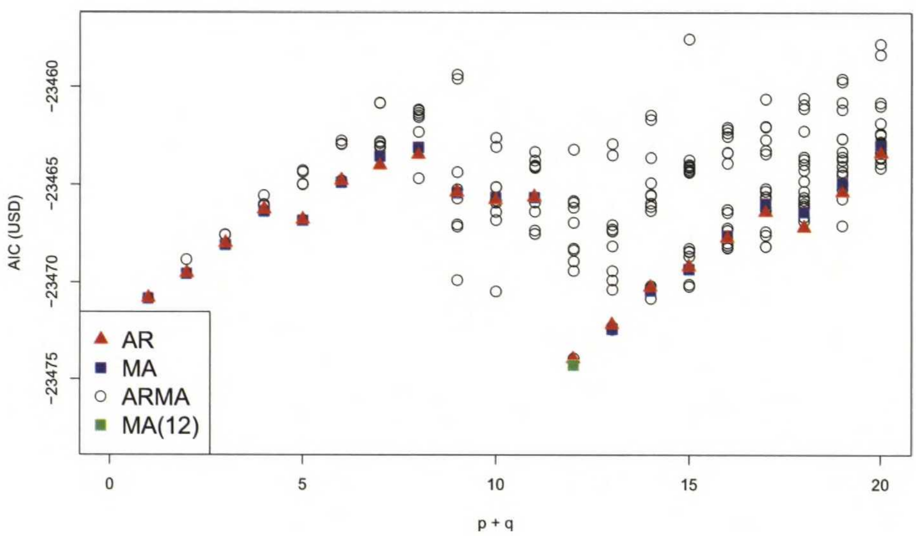

Figure 10: AIC values of some ARMA(p, q) models against $p + q$ for USD data

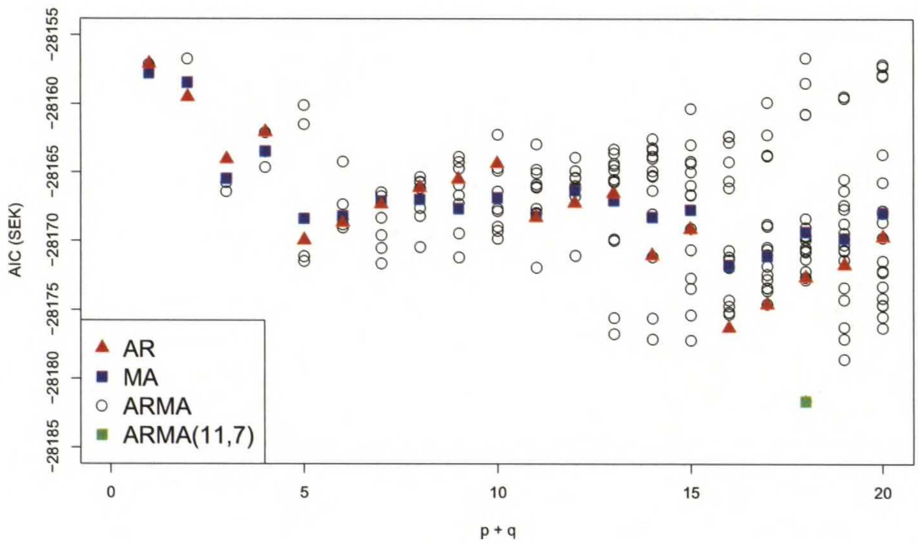

Figure 11: AIC values of some ARMA(p, q) models against $p + q$ for SEK data

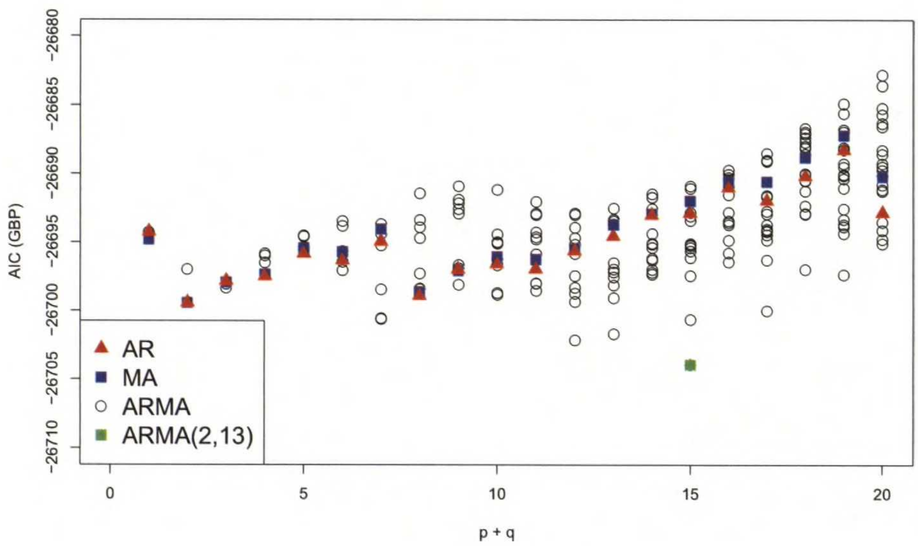

Figure 12: AIC values of some ARMA(p, q) models against $p + q$ for GBP data

According to the AIC values (Figure 10) the MA(12) model would be the best fit for the USD data, and the AR(12) has almost as low an AIC value. As for the SEK data (Figure 11) the AIC values suggest an ARMA(11,7) model and for the GBP data (Figure 12) an ARMA(2,13) model. When we compare the three AIC graphs, we note that for the SEK data the models with less than 5 parameters have relatively high AIC values as for USD and GBP data these simple models provide approximately as low AIC values as the models with more parameters. This could be caused by the non-symmetric QQ-plot and the large peaks at high lags in the ACF and PACF of the differenced logarithmic SEK series seen in Figure 8.

Another comparison method used is the Bayesian information criterion (BIC), which is similar to the AIC, but has a larger penalty on the number of parameters especially for longer time series (cf. Appendix A.5). The BIC values of the estimated ARMA models are shown in Figures 13, 14 and 15.

For the USD data, AR(1) and MA(1) models clearly have the lowest BIC values (Figure 13) and thus would be the best models for the data. The MA(12) model suggested by the AIC does not stand out at all. The BIC graphs for the SEK and the GBP series (Figure 14, Figure 15) are similar to that of the USD series. They also suggest AR(1) or MA(1) models and the ARMA models pointed out by the AIC do not have low BIC values.

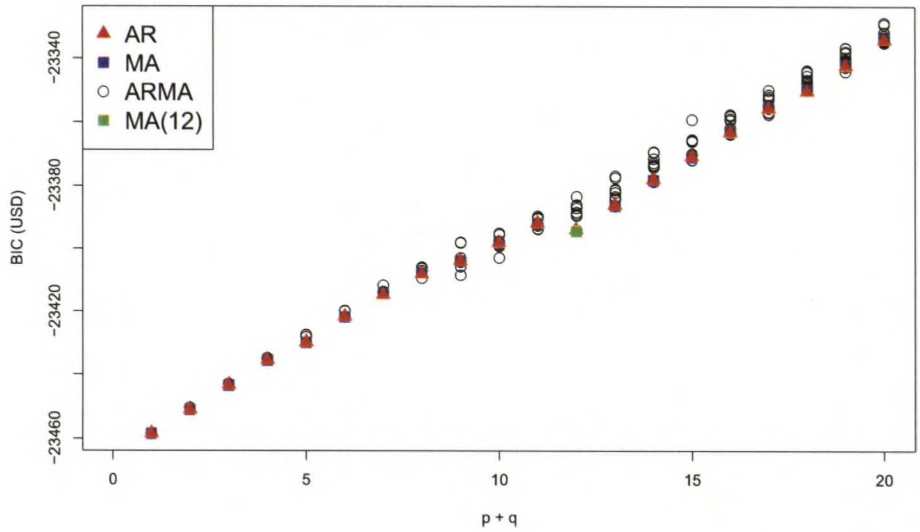

Figure 13: BIC values of some ARMA(p, q) models against $p+q$ for USD data

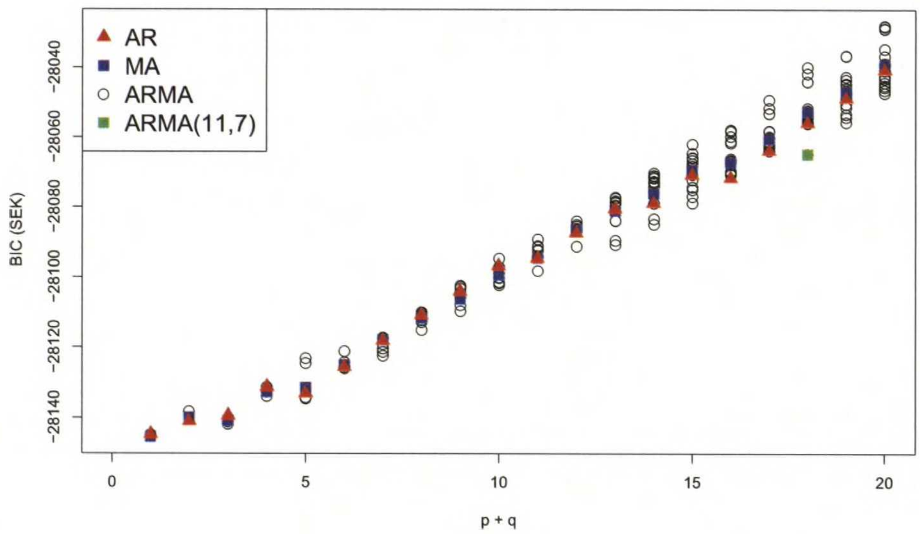

Figure 14: BIC values of some ARMA(p, q) models against $p+q$ for SEK data

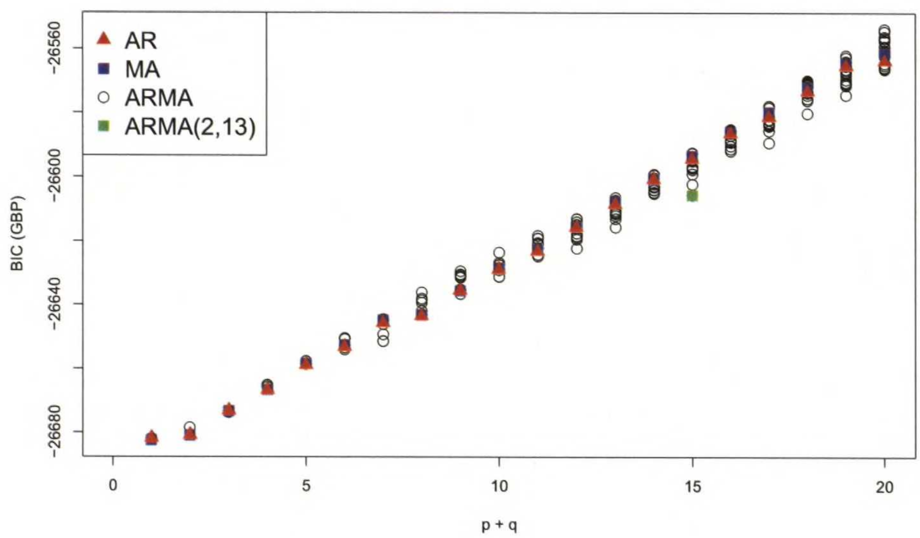

Figure 15: BIC values of some ARMA(p, q) models against $p + q$ for GBP data

In this case the AIC and the BIC provide different suggestions for the best ARMA models for all three datasets. However, we choose to follow the suggestion of the AIC since our goal is to reduce the serial correlations in the series, and following the BIC would mean to remove the serial correlation only at lag 1. In this setting, having more parameters is not considered to be such a disadvantage.

We discuss the U.S. dollar time series first. Since the AIC values pointed so clearly to an AR(12) or MA(12) model we now estimate these models for the differenced USD data and examine the residual series to evaluate the goodness of fit. The estimated coefficients for the AR(12) model are presented in Table 1, where the significant values are in bold. When comparing with the ACF and PACF of the differenced USD series in Figure 7 we see that the significant values of the coefficients are at the same lags where there were larger peaks in the ACF and PACF. Thus, the model seems sensible and should remove the serial correlations for lags smaller than 13.

Table 1: Estimated coefficients for AR(12) for USD data and their standard errors. The significant values are in bold.

coefficient	value	std. error	coefficient	value	std. error
ϕ_1	0.0045	0.0170	ϕ_7	0.0168	0.0170
ϕ_2	-0.0135	0.0170	ϕ_8	**0.0210**	0.0170
ϕ_3	0.0144	0.0170	ϕ_9	**-0.0345**	0.0170
ϕ_4	-0.0087	0.0170	ϕ_{10}	**-0.0254**	0.0170
ϕ_5	**0.0266**	0.0170	ϕ_{11}	**-0.0234**	0.0170
ϕ_6	0.0026	0.0170	ϕ_{12}	**0.0546**	0.0170
σ^2	6.561×10^{-5}				

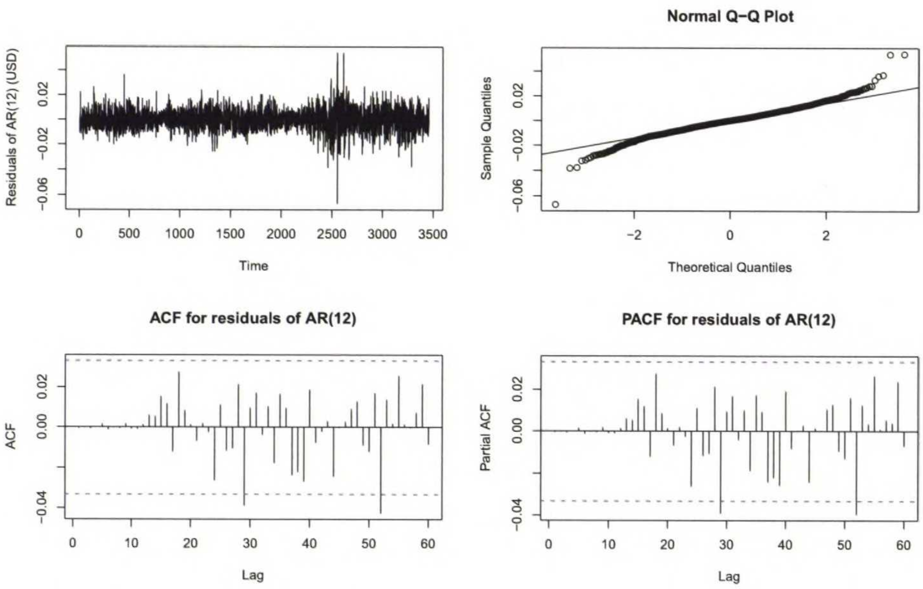

Figure 16: Residual analysis for the AR(12) model for USD data

Figure 16 presents the residuals of the AR(12) process applied to the differenced USD data, as well as the normal QQ-plot and the ACF and PACF. We can see that the AR(12) model removed all the correlations at lags smaller than 13, as it should have. However, there are still some serial correlations at larger lags. According to the QQ-plot the residuals are not normally distributed either. A Ljung-Box test for the correlations in the residuals (see Appendix A.4) provides a p-value that is very close to one at all degrees of freedom from one to 20. This suggests that, on all conventional significance levels, we have no evidence against the null hypothesis of the residuals being white noise. Thus, the AR(12) model was shown not to be a perfect fit to the data, but it did make it closer to white noise. We also estimated the MA(12) model for the differenced USD series and it gave the same significant coefficients and similar residuals as the AR(12).

Next, we tested how leaving out the insignificant coefficients (according to the ACF and PACF in Figure 7) in the AR model would change the model. We estimated an AR model for the differenced USD data such that coefficients other than ϕ_9 and ϕ_{12} were set to zero. The residual analysis of this model is shown in Figure 17 and the parameter values are in Table 2. The Ljung-Box test for the autocorrelation in the residuals gives a p-value between 0.5 and 1 for all degrees of freedom between one and 20. This provides no evidence against the null hypothesis on any conventional significance level, and thus, according to the test, the residuals can be considered to be white noise.

The QQ-plot in Figure 17 shows that the distribution of the residuals is symmetric but not normal. The autocorrelation functions of the two AR models for the USD data (AR(12) in Figure 16 and AR in Figure 17) have the same significant values left (29 and 52) but the AR(12) has 10 more (insignificant) parameters. Hence, we choose to work with the AR model with only two terms. It is simpler and still removes the significant serial correlation at lags 9 and 12.

Table 2: Estimated coefficients for the AR model for the USD data and their standard errors. Only the coefficients at lags 9 and 12 were estimated.

coefficient	value	std. error
ϕ_9	-0.0350	0.0170
ϕ_{12}	0.0552	0.0170
σ^2	6.581×10^{-5}	

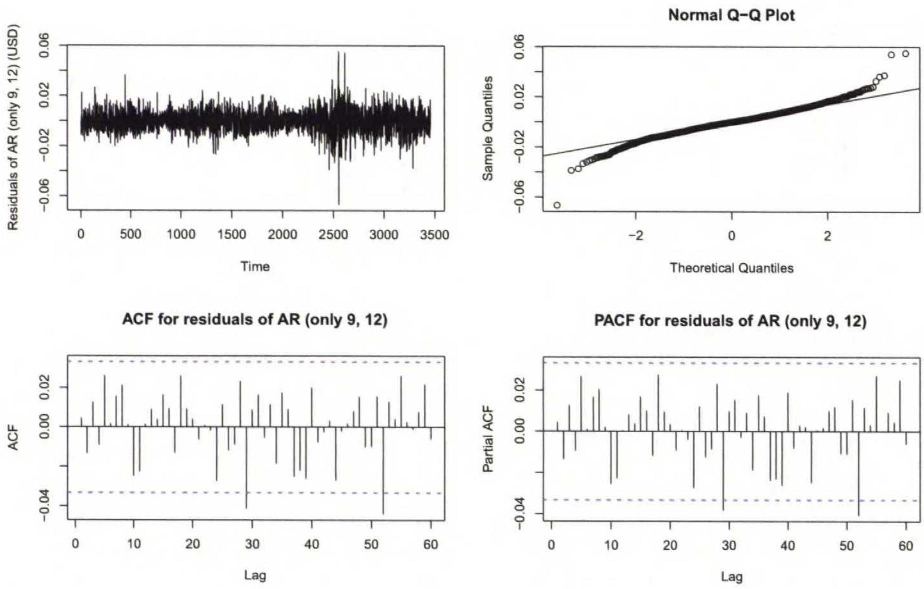

Figure 17: Residual analysis for the AR model with parameters only at lags 9 and 12 for USD data

In addition, we estimated an AR model for the USD series with only the coefficients ϕ_9, ϕ_{12} and ϕ_{29} as well as an AR model with coefficients ϕ_9, ϕ_{12}, ϕ_{29} and ϕ_{52}, corresponding to the significant peaks in the ACF in Figure 7. These models disposed of the peaks at lags 29 and 52, but did not otherwise have any effect on the ACF and PACF compared to the AR model with only two terms. Removing all the serial correlations is not considered to be important as 5% of the values in an ACF plot are expected to be outside the boundaries even though the series was white noise (on a 5% significance level). Hence, we continue working with the AR model with two terms. From now on we refer to this model by merely the AR model of the USD data.

Next we model the SEK series with the ARMA(11,7) model suggested by the AIC (Figure 11). The parameters of the ARMA(11,7) model are presented in Table 3. The significant values are in bold. The residual analysis of the ARMA(11,7) applied to the differenced logarithmic SEK series is shown in Figure 18. The QQ-plot shows that the residuals are not normally distributed, but they are fairly symmetric. The ACF and PACF show that there are still some serial correlations but they exceed the expected amount of 5% by very little. The Ljung-Box test for the autocorrelation in the residuals of the ARMA(11,7) model provides a p-value close to one for every lag until 30. This gives no evidence against the null hypothesis of white noise, on any conventional significance level.

Table 3: Estimated coefficients for the ARMA(11,7) model for the SEK data and their standard errors. Significant values are in bold.

coefficient	value	std. error	coefficient	value	std. error
ϕ_1	0.0121	0.0142	θ_1	**0.0424**	0.0026
ϕ_2	**-0.2844**	0.0155	θ_2	**0.2529**	0.0135
ϕ_3	**0.2880**	0.0161	θ_3	**-0.3232**	0.0125
ϕ_4	**0.1604**	0.0288	θ_4	**-0.1893**	0.0245
ϕ_5	**-0.6076**	0.0208	θ_5	**0.5463**	0.0066
ϕ_6	**-0.0734**	0.0294	θ_6	**0.1000**	0.0068
ϕ_7	**-0.4124**	0.0143	θ_7	**0.4098**	0.0062
ϕ_8	**0.0161**	0.0071			
ϕ_9	**-0.0190**	0.0164			
ϕ_{10}	**-0.0351**	0.0151			
ϕ_{11}	**0.0818**	0.0152	σ^2	1.679×10^{-5}	

Although the ARMA(11,7) was shown to be a good fit to the SEK data, it is quite complicated and has several parameters. It could be possible to find as good a model with less parameters by estimating an AR model and setting some parameters to zero. Since there were significant peaks in the ACF and PACF of the SEK data (Figure 8) on lags 1, 2, 3, 5, 11, 14, 16 and 21, we applied an AR model containing parameters corresponding to these. After applying models where not all of these 8 terms were included, we found that they were all necessary for removing the significant serial correlations. Thus, we decided on an AR model with only the parameters ϕ_1, ϕ_2, ϕ_3, ϕ_5, ϕ_{11}, ϕ_{14}, ϕ_{16} and ϕ_{21}. The estimated coefficients are presented in Table 4 and the residual analysis in Figure 19.

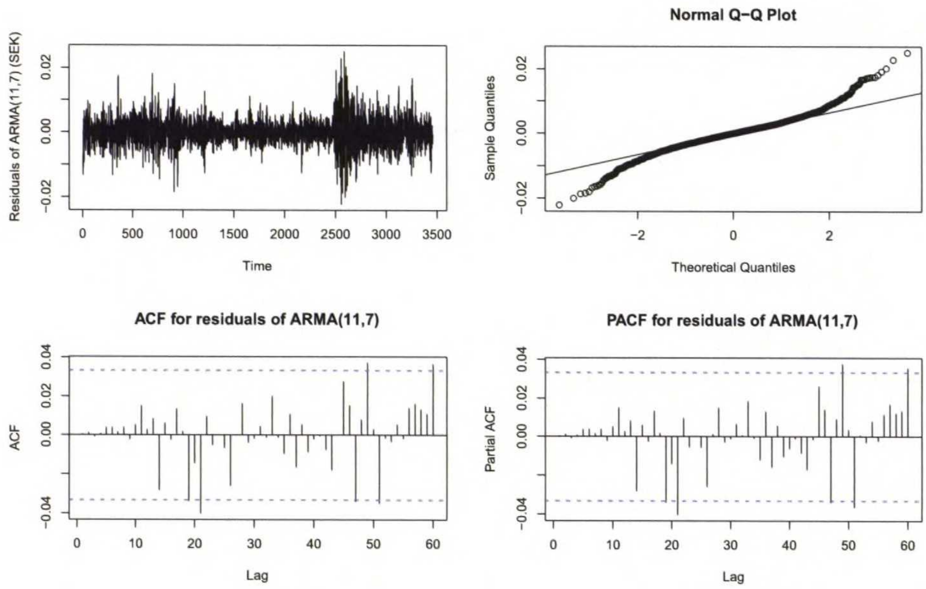

Figure 18: Residual analysis for the ARMA(11,7) model for SEK data

Table 4: Estimated coefficients for the AR model for the SEK data and their standard errors. Only the coefficients at lags 1, 2, 3, 5, 11, 14, 16 and 21 were estimated.

coefficient	value	std. error
ϕ_1	0.0544	0.0169
ϕ_2	-0.0358	0.0169
ϕ_3	-0.0475	0.0170
ϕ_5	-0.0501	0.0170
ϕ_{11}	-0.0384	0.0169
ϕ_{14}	-0.0390	0.0170
ϕ_{16}	0.0459	0.0170
ϕ_{21}	-0.0647	0.0170
σ^2	1.678×10^{-5}	

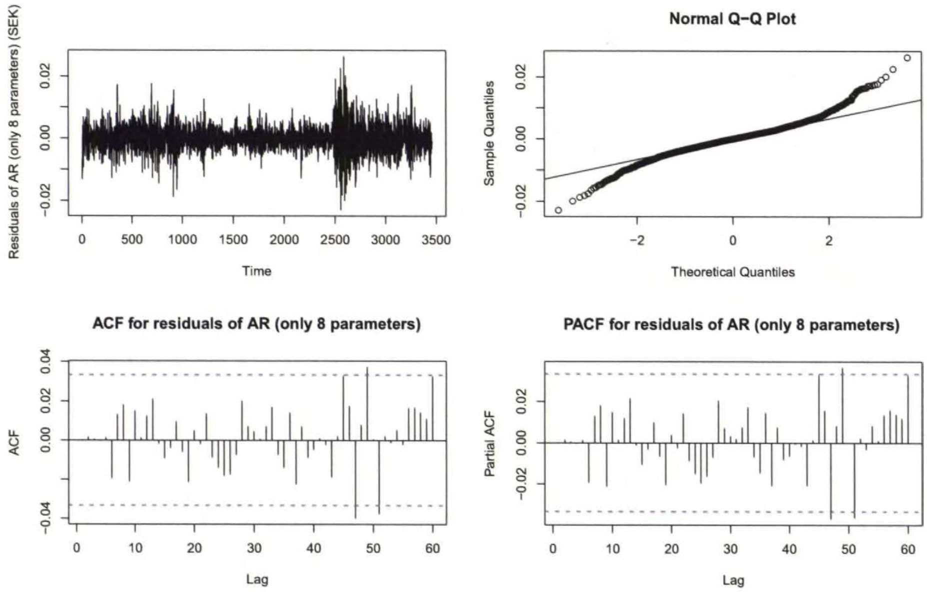

Figure 19: Residual analysis for the AR model with parameters only at lags 1, 2, 3, 5, 11, 14, 16 and 21 for SEK data

Comparing this model with the ARMA(11,7) we see that the AR model with only 8 parameters is at least as good a fit to the SEK data. The distribution of the residuals is not normal but quite symmetric. The AR model removed all the significant serial correlations until lag 48 and especially disposed of the peak at lag 21, which was still visible in the ACF of the ARMA(11,7) model. The ACF and PACF (Figure 19) indicate also that there are no more serial correlations since only 5% of the peaks reach outside the boundaries. The Ljung-Box test for the correlations in the residuals provides a p-value close to one for all lags up to 30. Thus, on all conventional significance levels there are no evidence against the null hypothesis of the residuals being white noise. We conclude that the AR model with the 8 parameters is the best choice for modelling the serial correlations in the SEK data.

Finally, we proceed to the GBP series. The AIC (Figure 12) suggests an ARMA(2,13) model hence we first estimate it for the differenced logarithmic GBP series. The estimated coefficients are presented in Table 5. The significant values are in bold. Since many of the parameters were not significant (especially at higher lags), it could be sensible to estimate an AR model with some parameters set to zero.

Table 5: Estimated coefficients for the ARMA(2,13) model for the GBP data and their standard errors. Significant values are in bold.

coefficient	value	std. error	coefficient	value	std. error
ϕ_1	**0.1488**	0.0168	θ_7	-0.0010	0.0229
ϕ_2	**-0.9421**	0.0169	θ_8	-0.0218	0.0221
			θ_9	-0.0092	0.0226
θ_1	**-0.1071**	0.0224	θ_{10}	**-0.0575**	0.0177
θ_2	**0.8952**	0.0223	θ_{11}	**-0.0317**	0.0197
θ_3	**0.0335**	0.0225	θ_{12}	0.0100	0.0183
θ_4	**-0.0612**	0.0226	θ_{13}	**-0.0319**	0.0198
θ_5	-0.0009	0.0226			
θ_6	0.0037	0.0229	σ^2	2.575×10^{-5}	

The residual analysis of the ARMA(2,13) model for the GBP data is shown in Figure 20. The QQ-plot shows that the residuals are not normally distributed but are almost symmetric. The ACF and PACF indicate that there are no serial correlations at lags smaller than 21, but some peaks at higher lags remain. The Ljung-Box test for the autocorrelation in the residuals provides a p-value close to one for all lags until 30. Thus we have no evidence against the null hypothesis on any conventional significance level, and the residuals are white noise.

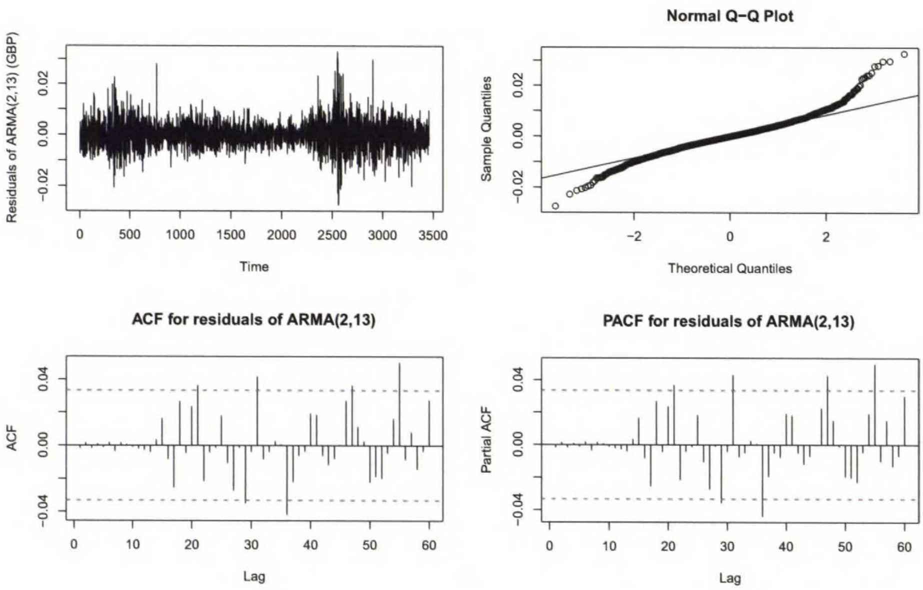

Figure 20: Residual analysis for the ARMA(2,13) model for GBP data

To find out whether an AR model with reduced parameters would be a better choice for the GBP data we estimate an AR model with only parameters ϕ_1, ϕ_2 and ϕ_8. We choose these parameters since there are peaks in the ACF and PACF in Figure 9 at these lags and the next peak is not until lag 20. The estimated coefficients of the AR model are shown in Table 6. The residual analysis of the AR model with only the parameters ϕ_1, ϕ_2 and ϕ_8 is presented in Figure 21. The residuals are not normally distributed but are almost symmetric. The ACF and PACF show that there are still some serial correlations in the residuals but not more than in the residual of the ARMA(2,13) model (Figure 20). The Ljung-Box test for the autocorrelation gives a p-value larger than 0.5 for all lags until 30 which provides no evidence against the null hypothesis of white noise, on any conventional significance level. This AR model with only three parameters was shown to be as good a fit to the GBP data as the ARMA(2,13), and since the AR model is less complicated we choose it to model the serial correlations. From now on we refer to this AR model with only three terms as solely the AR model of the GBP series.

Table 6: Estimated coefficients for the AR model for the GBP data and their standard errors. Only the coefficients at lags 1, 2 and 8 were estimated.

coefficient	value	std. error
ϕ_1	0.0413	0.0170
ϕ_2	-0.0442	0.0170
ϕ_8	-0.0428	0.0170
σ^2	2.593×10^{-5}	

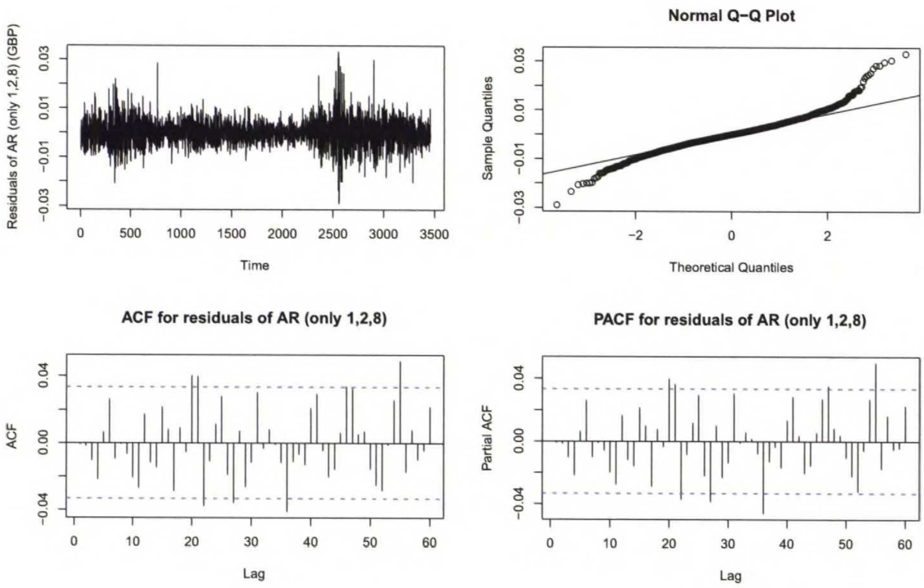

Figure 21: Residual analysis for the AR model with parameters only at lags 1, 2 and 8 for GBP data

When we look at the residual series for the chosen AR models in Figures 17, 19 and 21 we still detect some volatility clustering, i.e., the higher variances occur together. This kind of behavior can be expected to be better captured by applying a heteroskedastic model such as ARCH or GARCH (Appendix A.7).

3.4 Testing for heteroskedasticity

Next we proceed to model the changing volatility in the time series. As mentioned earlier, this can be modelled with models of heteroskedasticity (see Appendix A.7). To more formally justify the usage of heteroskedastic models we calculate some simple statistical measures which show whether the variances of the three time series actually change over time. First we examine the ten-point moving standard deviations of the differenced USD series and the differenced logarithmic SEK and GBP series presented in Figures 22, 23 and 24, respectively. The graphs clearly shows that the standard deviation varies significantly at different time periods. We calculated also the 3-point, 5-point, and 20-point moving standard deviations, but the ten-point versions showed best the variations in the standard deviation.

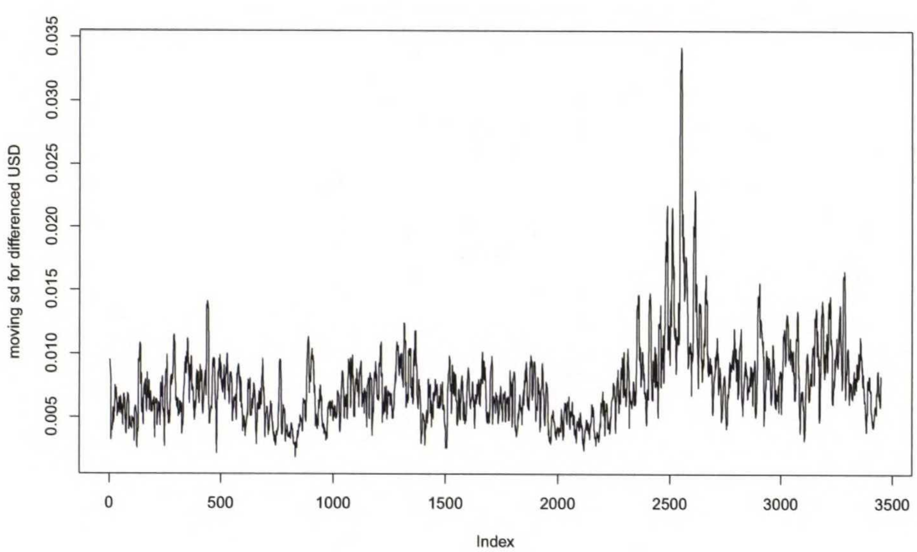

Figure 22: Ten-point moving standard deviation for the differenced USD series

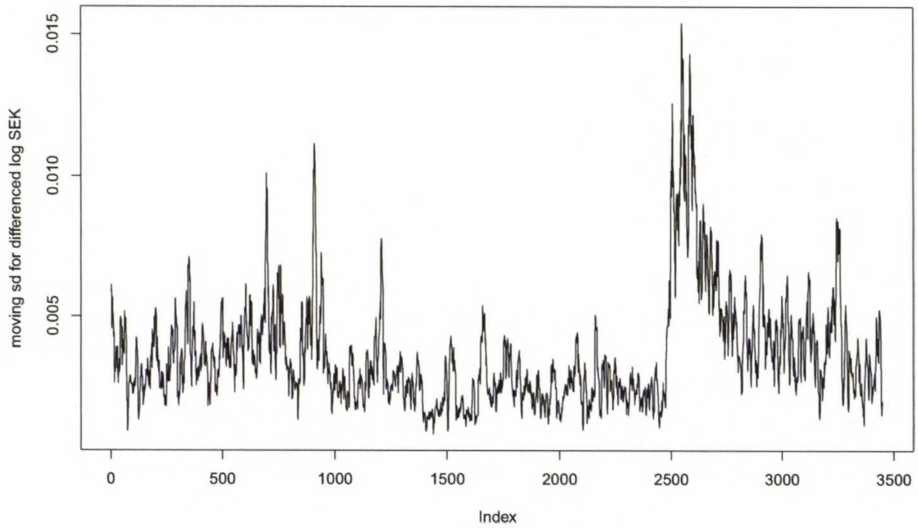

Figure 23: Ten-point moving standard deviation for the differenced logarithmic SEK series

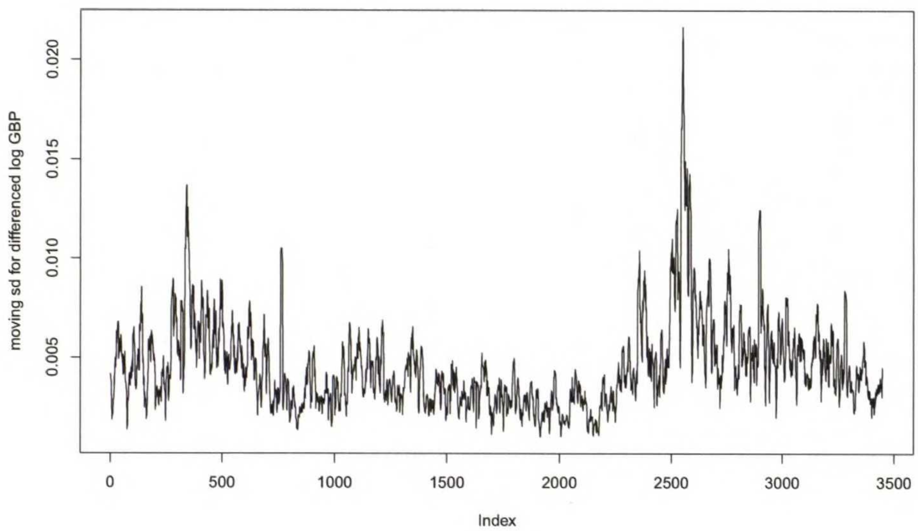

Figure 24: Ten-point moving standard deviation for the differenced logarithmic GBP series

Next we use the coefficient of variation c_v defined as the ratio of the standard deviation σ to the mean μ

$$c_v = \frac{\sigma}{\mu}. \tag{1}$$

The coefficient of variation is a statistical measure which allows to compare how wide spread two probability distributions are, or how large is the change in volatility within one time series [Searls, 1964].

The ten-point coefficient of variation for the three time series are shown in Figures 25, 26 and 27. From the figures it is quite clear that the series have heteroskedastic characteristics since there are several large (positive and negative) values. We again calculated also the 3-point, 5-point, and 20-point coefficients of variation for the three series, but they were not any clearer in detecting the heteroskedasticity than the ten-point versions.

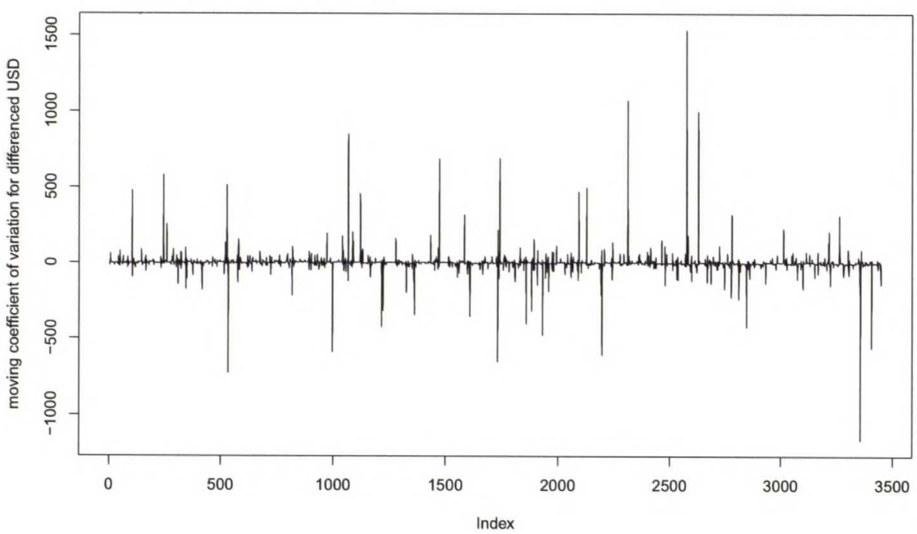

Figure 25: Ten-point moving coefficient of variation for the differenced USD series

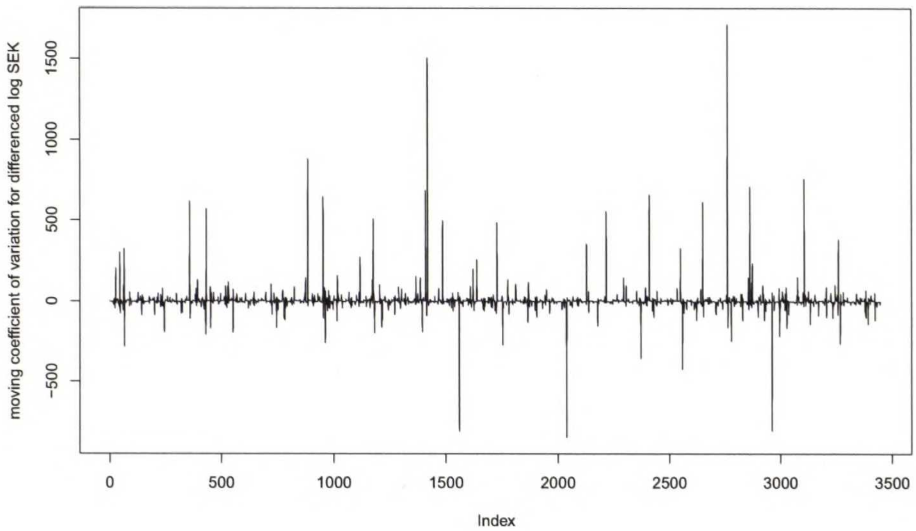

Figure 26: Ten-point moving coefficient of variation for the differenced logarithmic SEK series

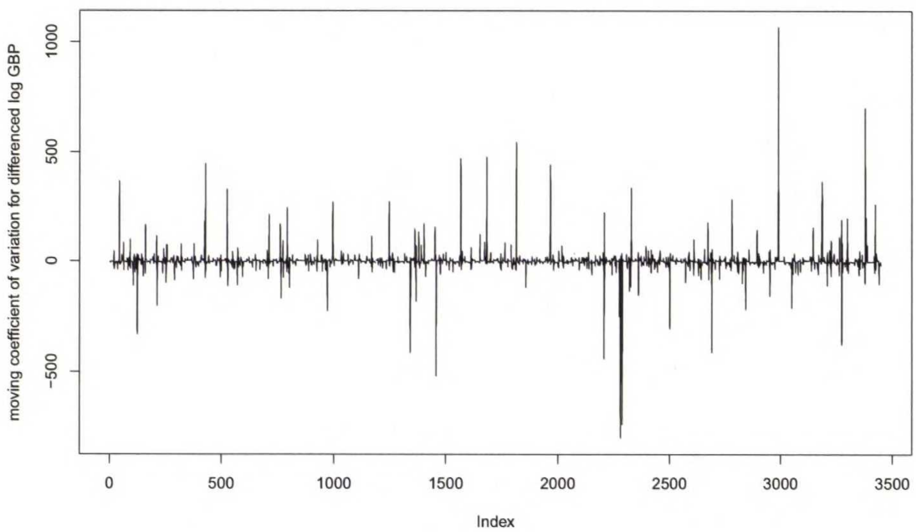

Figure 27: Ten-point moving coefficient of variation for the differenced logritmic GBP series

In Figure 28 are the correlogram and partial correlogram for the squared differenced USD series and the squared differenced logarithmic SEK and GBP series (Appendix A.2). The graph shows that there are a lot of serial correlations in the squared series which also points strongly to a heteroskedastic model.

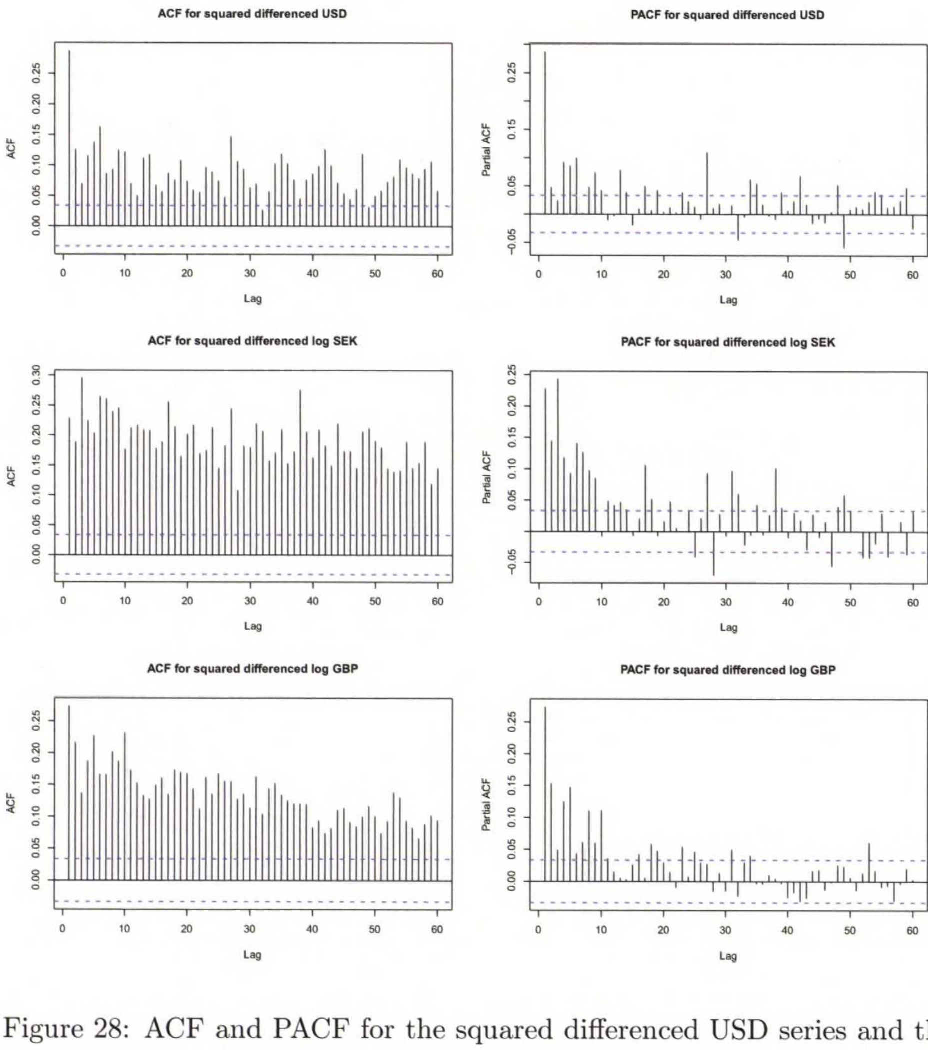

Figure 28: ACF and PACF for the squared differenced USD series and the squared differenced logarithmic SEK and GBP series

Next we use the Lagrange multiplier method (see Appendix A.9) for deriving an ARCH test. This test can confirm whether the heteroskedasticity that could be seen in the residuals of the AR models in Figures 17, 19 and 21 is statistically significant. The test gives a p-value very close to zero at all lags between 1 and 30 for all three time series, and thus, on all conventional significance levels, we have significant evidence against the null hypothesis of no conditional correlations in the variances. This means that an ARCH or a GARCH model should be a fairly good fit to the datasets.

Still another method to detect heteroskedasticity is to apply the Ljung-Box test (Appendix A.4) to the squared time series [Tsay, 2010]. The possible autocorrelation in the squared series would mean that we need a model of heteroskedasticity to properly model the series. The p-value of the Ljung-Box test statistic, for all three datasets, is close to zero suggesting that the variances really are serially correlated.

3.5 Estimating ARCH and GARCH models

We start the estimation of ARCH and GARCH processes (Appendix A.7) with the simplest models to see if they give satisfying results. We will apply ARCH(1), ARCH(2) and GARCH(1,1) models to the three residual series of the AR processes chosen in Section 3.3. The GARCH(1,1) is usually accurate enough to model the volatility clustering in a time series. And an ARCH(m) with $m > 2$ would not be sensible to estimate since it has more parameters than GARCH(1,1).

From a theoretical viewpoint we can argue that a GARCH model should be a better fit for the residual data since it allows the model to depend also on the past estimated variance and not just the realized variances in the residual series (as in ARCH models). This should make it possible for the GARCH models to be more accurate than the plain ARCH models. There are also more complex ARCH models (Appendix A.7.3) which can in some cases be significantly better for the data. However, here we do not detect unsymmetric reactions to positive and negative peaks or any other exceptional behavior which would require the use of these models.

It should be possible to determine the degree of an ARCH(m) process by looking at the PACF of the squared series (Figure 28) [Bollerslev, 1986]. If the partial correlogram clearly cuts off after lag h, then an ARCH(h) model could be a good fit to the data. In Figure 28 the PACF graphs of the squared SEK and the squared GBP series do not have clear cut off points, but they die out slowly and there are significant peaks here and there also at higher lags. This suggests that a plain ARCH process cannot model these two datasets as well as a GARCH process. However, in the PACF of the squared USD series the value at lag 1 is significantly larger than the other peaks. This points out that an ARCH(1) process could be a good fit to the USD data.

We start with applying an ARCH(1) model to the residuals of the AR model for the differenced USD data. Figure 29 presents the residuals of the ARCH(1) process, a QQ-plot against normal distribution and the ACF and PACF. From Figure 29 we see that there are no more significant serial correlations, and that the residual series is not normally distributed. The residuals seem to still have some volatility clustering left, even though the PACF of the squared series suggested that an ARCH(1) model could be enough.

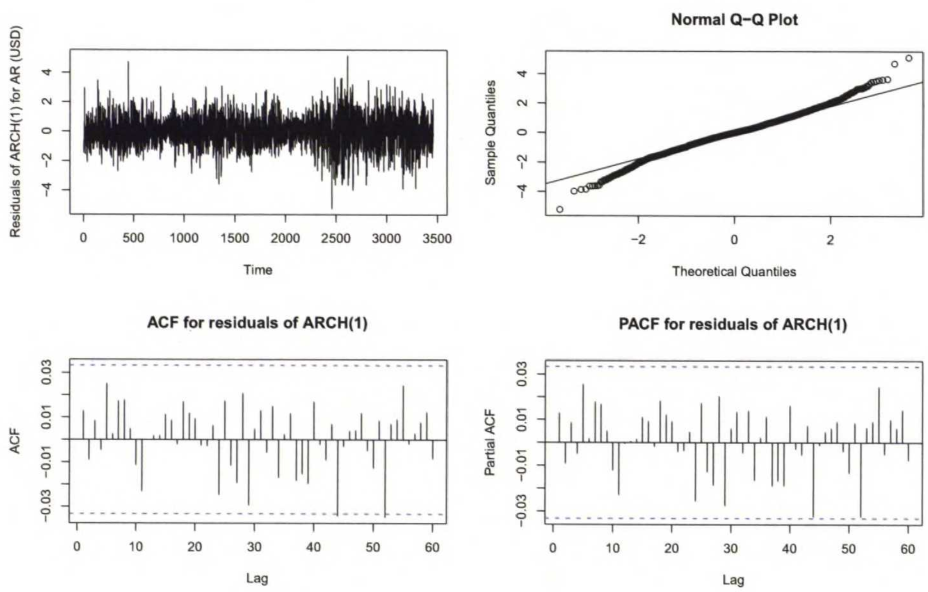

Figure 29: Residual analysis for ARCH(1) for the residuals of the AR model of USD data

The QQ-plot in Figure 29 shows that the distribution of the residual series has heavier tails than the normal distribution. The distribution is also quite symmetric. Student's t-distribution satisfies these conditions and thus we examine if the residuals would follow a t-distribution. In Figure 30 is a QQ-plot of the residuals against t-distribution with 10 degrees of freedom. We tried several different degrees of freedom for the t-distribution, but the one with 10 degrees of freedom seemed to be the best fit to the distribution of the residuals. We can see that the residual series follows quite well the t-distribution.

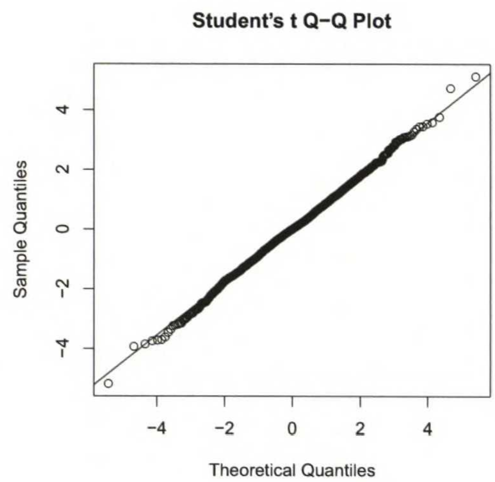

Figure 30: QQ-plot for the ARCH(1) residuals of the AR residuals of the USD data against t-distribution with 10 degrees of freedom

Although ARCH(1) was shown to be a fairly good model, we estimate also the ARCH(2) and the GARCH(1,1) model for the residual of the AR model of the USD data to see if the results differ. Especially to see if we can remove the remaining volatility clustering. The GACH(1,1) was a better model to dispose of the volatility clustering than the ARCH(2). The residuals of GARCH(1,1), a QQ-plot against t-distribution with 10 degrees of freedom and the ACF and PACF are presented in Figure 31. According to the residual analysis the GARCH(1,1) applied to the residuals of the AR model is also a good fit to the data. There are no serial correlations, the residuals appear to be close to white noise and they are t-distributed with 10 degrees of freedom.

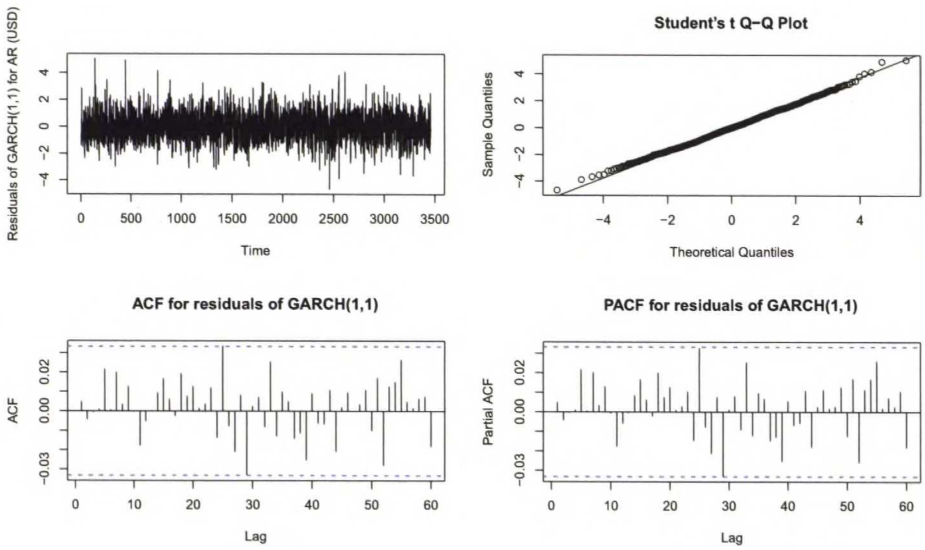

Figure 31: Residual analysis for GARCH(1,1) for the residuals of the AR model of USD data. QQ-plot against t-distribution with 10 degrees of freedom.

When we compare the residual analysis of the ARCH(1) and the GARCH(1,1) for the USD data (Figures 29 and 31) we see a few differences. The residual series of ARCH(1) seems to still have some volatility clustering, but for GARCH(1,1) the series is more stable in terms of variance. Then again, the ACF and PACF show that the serial correlations (even though not significant) are a little larger for the GARCH(1,1) process. All in all, GARCH(1,1) might be a better model, since it removed the volatility clustering which was the aim of applying the models of heteroskedasticity.

Next we applied ARCH(1), ARCH(2) and GARCH(1,1) models to the residuals of the AR model for the differenced logarithmic SEK series. As happened with the USD data, the ARCH(1) and ARCH(2) models did not remove the volatility clustering as well as the GARCH(1,1) model. The residual analysis of the GARCH(1,1) applied to the AR of the SEK data is shown in Figure 32. The QQ-plot is against t-distribution with 10 degrees of freedom which was again the best fit. There are no more significant serial correlations. The few peaks that reach outside the boundaries in the ACF and PACF are considered to be meaningless since 5% of the values are expected to be outside the boundaries even if the series is white noise. Also, there is no more visible volatility clustering in the residual series.

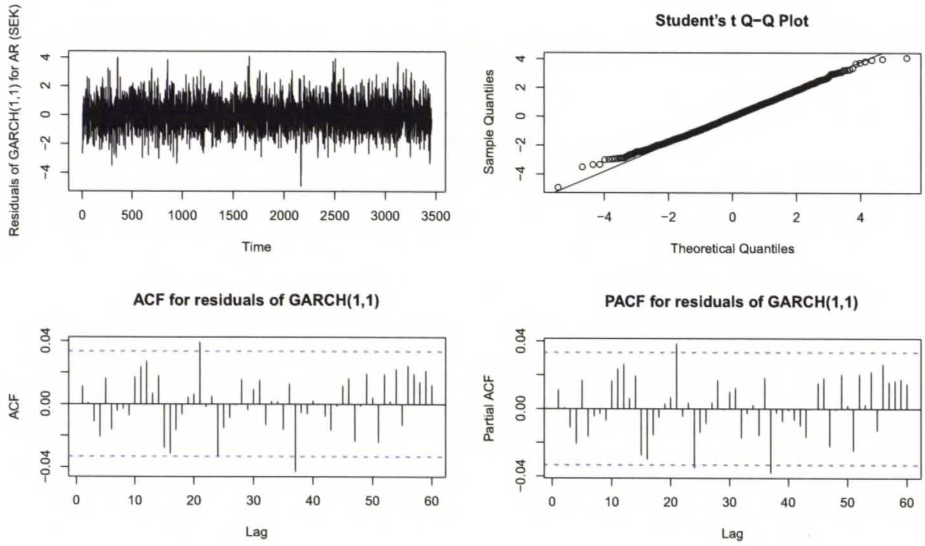

Figure 32: Residual analysis for GARCH(1,1) for the residuals of the AR model for SEK data. QQ-plot against t-distribution with 10 degrees of freedom.

Finally we estimated the ARCH(1), ARCH(2) and GARCH(1,1) models to the residuals of the AR model of the differenced logarithmic GBP series. The GARCH(1,1) is again clearly the best model and its residual analysis is presented in Figure 33. The QQ-plot is again against t-distribution with 10 degrees of freedom, which was the best fit to the series. There are still some serial correlations according to the ACF and PACF but as was already pointed out, it is not always necessary to remove all the serial correlations at the expense of adding more parameters. The residual series seems to be very close to white noise and the volatility clustering is no more significant.

All in all, the GARCH(1,1) was the best fit to all three residual series and it removed the heteroskedastic effects quite efficiently. Thus, we have succeeded to model both the serial correlations (with ARMA models) and the volatility clustering (with models of heteroskedasticity) in the three exchange rate series.

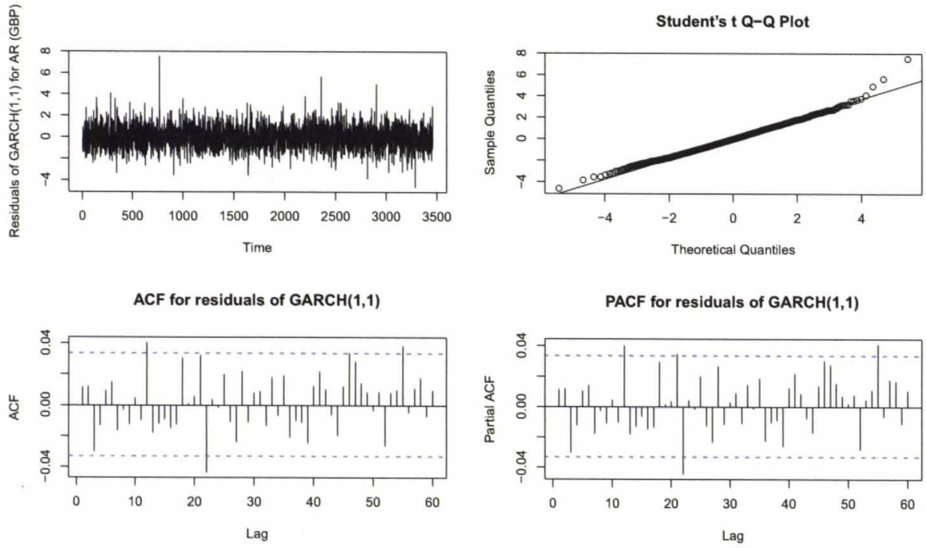

Figure 33: Residual analysis for GARCH(1,1) for the residuals of the AR model for GBP data. QQ-plot against t-distribution with 10 degrees of freedom.

3.6 Dividing the data

Even though we concluded that the ARMA and GARCH models estimated in previous sections are a good fit to the datasets, we still look into the possibility of finding a better model by dividing the data into two parts. When we look at the residual series of the AR model of the USD data (Figure 34) we can detect two different parts of the data (separated by the red vertical line). During the first part (from the beginning of the data until the end of 2007) there is relatively less volatility than during the second part (from the beginning of 2008 until the end of the data). The increased volatility after the year 2008 is likely to be due to the worldwide financial crisis which made investments in foreign currency riskier than before. Now we will estimate some ARCH and GARCH models separately to the two parts to see if this improves the goodness of fit for the heteroskedastic models compared to the GARCH(1,1) estimated earlier.

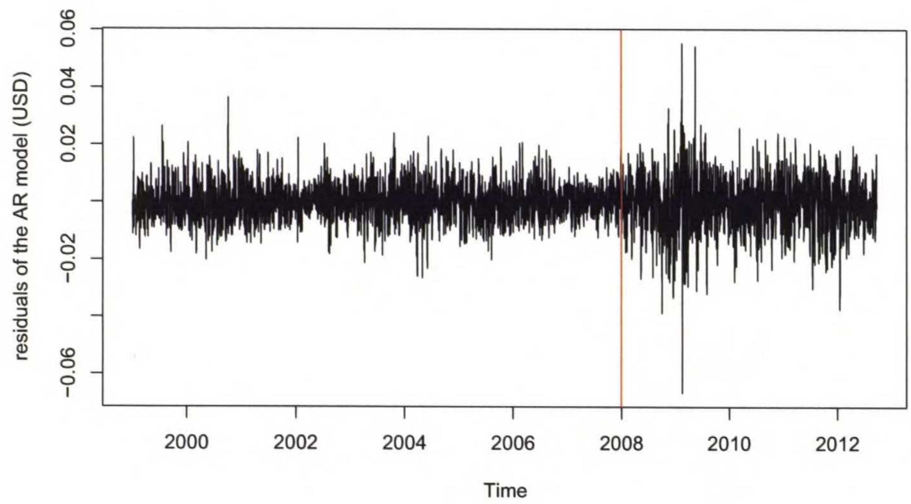

Figure 34: The residual series of the AR model for USD data. The data is divided at the red line.

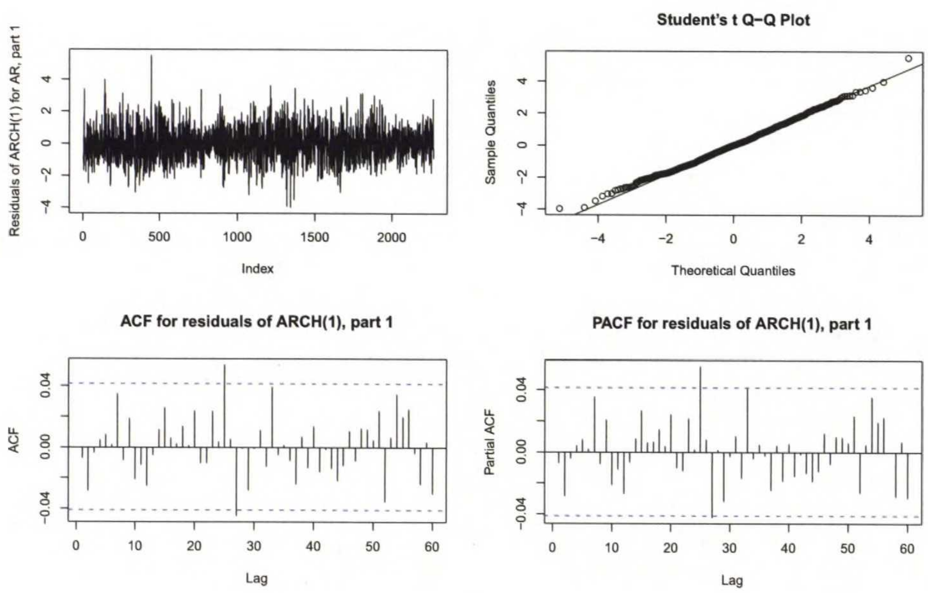

Figure 35: Residual analysis for ARCH(1) for residuals of the AR model of USD data. QQ-plot against t-distribution with 10 degrees of freedom. First part of the data.

We estimate ARCH(1), ARCH(2) and GARCH(1,1) for both parts separately. For the first part, the three models showed very similar results in terms of residual analysis and Ljung-Box test. Hence, we would choose the simplest model, i.e., ARCH(1) for the first part. The residual analysis for ARCH(1) applied to the residuals of the AR model of USD data is shown in Figure 35. The QQ-plot is constructed against t-distribution with 10 degrees of freedom. We tried several different degrees of freedom, but the distribution with 10 degrees of freedom seemed to be the best fit. The Ljung-Box test for the autocorrelation in the residuals of ARCH(1) gives a p-value of 0.734 at lag 1, and stays above 0.3 for all other lags. This suggests that the residuals actually can be considered to be white noise.

When we compare the residual analysis of ARCH(1) for the first part and GARCH(1,1) for the whole data (Figure 31) we see that there is more volatility clustering in the first part residuals. In addition, the ACF and PACF show that there are more serial correlations in the first part. This suggests that maybe the AR model is not good for the first part of the data but was chosen because it modelled well the correlations in the second part of the USD data.

For the second part the GARCH(1,1) model made the residuals closer to white noise than the plain ARCH models and thus we choose GARCH(1,1) for the second part of the data. The residual analysis of GARCH(1,1) applied to the AR model residuals for the second part of USD data is presented in Figure 36. The Ljung-Box test for the autocorrelation in the residuals of GARCH(1,1) gives a p-value of 0.294 at lag 1 and a p-value larger than 0.5 for other lags. This means that, on all conventional significance levels, the residuals can be considered to be white noise.

Compared to the GARCH(1,1) for the whole data, the residual series of GARCH(1,1) for the second part is quite similar. There are no serial correlations in either of these series. The residuals of the GARCH(1,1) for the second part are t-distributed with 10 degrees of freedom, as were the residuals for the whole data.

We could say that there is more heteroskedasticity in the second part of the USD data than the first part since a GARCH model was needed for the second part as for the first part an ARCH model was enough. However, applying a GARCH model also to the first part can only make the model more accurate and thus there is no reason not to use the GARCH(1,1) model for the whole data, as we did earlier. It is also more convenient and less complicated to use one model instead of two different models, and thus

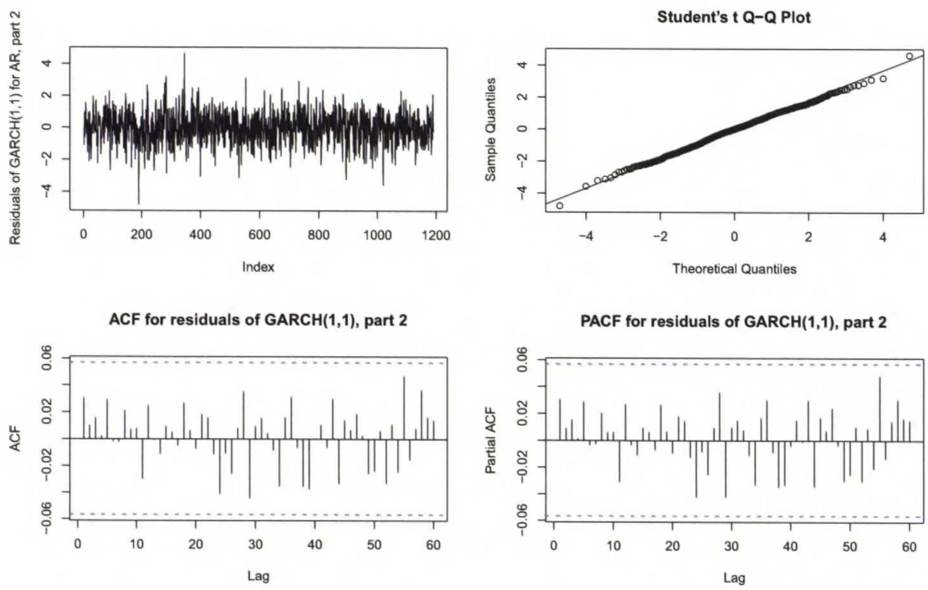

Figure 36: Residual analysis for GARCH(1,1) for residuals of the AR model of USD data. QQ-plot against t-distribution with 10 degrees of freedom. Second part of the data.

we conclude that the GARCH(1,1) applied to the whole residual series of the AR model is the best fit to the USD data.

The SEK and the GBP series were quite similar to the USD data and it seems that dividing them would not provide better goodness of fit since this was the case for the USD data. In Figures 8 and 9 the differenced logarithmic series of SEK and GBP data do not seem to even have clearly two different parts since there are more volatility also in the beginnings of these two datasets compared to the USD data. Thus, we might have to divide the SEK and the GBP series to at least three parts which would make the model significantly more complex. In addition, the models found earlier were already sufficiently accurate which makes dividing the data unnecessary.

4 Conclusions

In this thesis we took three exchange rate datasets, namely the United States dollar (USD), the Swedish krona (SEK), and the pound sterling (GBP), each against the euro, and studied them from the point of view of time series analysis. The aim was to find time series models which explained the behavior of the series as accurately as possible while keeping the number of parameters in the models at a sensible level. The three time series were quite similar but had also some differences. Especially the behavior of the SEK data after the financial crisis in 2008 was distinctively different from the other two series.

The modelling included two parts: first we removed the serial correlations using autoregressive moving average (ARMA) models and then applied models of heteroskedasticity to the residuals of the ARMA models to eliminate the volatility clustering. The serial correlations were caused by the fact that the rate of today usually depends somehow on the past values. The volatility clustering, meaning that higher variances occur together, can be explained by normal human behavior. At uncertain times the rate varies greatly as at more stable times the rate hardly fluctuates at all.

Before starting the actual modelling of the datasets we discussed the role of volatility in economic time series. Volatility here means the non-constant fluctuations around a constant mean or a trend line and is of particular interest since it is closely related to risk. Risk in the exchange market, as on any economic market, is the factor which along with the previous prices determines the current price of the asset. Thus understanding and modelling changes in volatility is crucial.

The phenomenon of volatility clustering and time-dependent variances is called heteroskedastic effect. This property of the data can be modelled with models of heteroskedasticity, the most common of which are autoregressive conditional heteroskedasticity (ARCH) and generalized autoregressive conditional heteroskedasticity (GARCH) models (introduced by [Engle, 1982] and [Bollerslev, 1986]).

The analysis of the three exchange rate data was started by choosing a suitable ARMA model for each of the datasets. We used the Akaike information criterion and the Bayesian information criterion to determine the number of parameters needed for the model to sufficiently remove the serial correlations. We applied several models to each data and used graphical residual analysis containing normal QQ-plots, correlograms and partial correlograms to finally decide on the best models. The chosen ARMA models were an AR model with only parameters ϕ_9 and ϕ_{12} for the USD data, an AR model with parameters ϕ_1, ϕ_2, ϕ_3, ϕ_5, ϕ_{11}, ϕ_{14}, ϕ_{16} and ϕ_{21} for the SEK data and an AR model with parameters ϕ_1, ϕ_2 and ϕ_8 for the GBP data. These models removed all the significant serial correlations in all three time series.

In the residuals of the chosen ARMA models we could detect some volatility clustering which pointed to the necessity of models of heteroskedasticity. However, we also tested more formally whether there were heteroskedastic effects in the data. We studied the moving standard deviations and moving coefficients of variation and they indicated that all three series had heteroskedasticity in them. We also applied an ARCH test based on the Lagrange multipliers and it pointed clearly to heteroskedasticity in the three datasets.

Next we applied some simple models of heteroskedasticity to the residuals of the three AR models. We estimated the ARCH(1), ARCH(2) and GARCH(1,1) models to each residual series. By looking at the residual analyses of these models we concluded that the ARCH(1) and ARCH(2) could not remove the volatility clustering in any of the datasets but the GARCH(1,1) was a good fit for all three time series. It disposed of the volatility clustering and there were no serial correlations in the residual series. We also detected that the residuals followed closely a t-distribution with 10 degrees of freedom.

Finally, since there were two distinctively differently behaving parts in the USD data, we studied whether modelling these two parts separately would give even better results. However, modelling the two parts individually did not provide significantly better goodness of fit than having one model for the whole USD data. We concluded that this would probably be the case for the SEK and the GBP data also, since they were similar to the USD data.

All in all we were satisfied with the models we found for each dataset. They described well the behavior of the three exchange rate datasets modelling both the serial correlations and the heteroskedasticity in the data. The models were also relatively simple and did not have too many parameters to be sensible in practice.

While the results of this study were quite satisfying even more accurate models could be achieved by using more complex ARCH models (Appendix A.7.3). Thus, the next step would be applying several other ARCH based models and comparing the results with the residual analyses and other indicators of goodness of fit of this study. Probably these models would not give significantly better results than the GARCH(1,1) model, since it was already a good fit, but checking this could be useful.

Another approach for continuing with the theme of this thesis would be to take other economic datasets, model them and compare the results with this study. It would be interesting to see what kind of similarities and differences can be found. This could give some deeper insight into how exchange rates and other economic time series (e.g., stock share prices) differ. In the exchange rate series we found no seasonality of any kind, but we cannot say if this is merely a coincidence and there might be some seasonality in other time series of exchange rates or other economic series. The existence of, e.g., a 5-day seasonality would be sensible since the data are usually only for weekdays and the events during the weekends could have a larger effect on the rates. Also, comparing the heteroskedastic effects in the data would be interesting.

A third possible way to extend this study would be to apply multivariate GARCH models to a set of exchange rates [Engle and Kroner, 1995]. These models combine the heteroskedasticity of several time series and model them together. Thus, if there is some dependence in the volatility clustering of two or more series these dependencies could be taken into account.

A Theory

A.1 White noise

A time series with independent, identically distributed random variables is called white noise. The process has got its name from the fact that it has all frequencies equally present, as in white light. A time series $\{y_t\}$ is a Gaussian white noise process if

$$y_t = \varepsilon_t, \tag{2}$$

where $\{\varepsilon_t\}$ is a sequence of independent random variables which follow a normal distribution with mean zero and variance σ^2. [Davison, 2011]

A.2 ACF and PACF

Empirical autocorrelation function ACF (or correlogram) of a time series is a useful tool for examining the serial correlations. To calculate the empirical autocorrelation function we first need the empirical autocovariance function of equally spaced data $y_1, \ldots, y_n$, which is

$$\hat{\gamma}_h = \frac{1}{n-h-1} \sum_{i=1}^{n-h} (y_i - \overline{y})(y_{i+h} - \overline{y}), \quad h = 0, 1, \ldots, n-2, \tag{3}$$

where $\overline{y}$ is the average of $y_1, \ldots, y_n$, i.e., $\overline{y} = \frac{1}{n} \sum_{i=1}^{n} y_i$. The empirical variance of the data $y_1, \ldots, y_n$ is $\hat{\gamma}_0 = \frac{1}{n-1}(y_i - \overline{y})^2$. The correlogram is then a plot of the empirical autocorrelation function $\hat{\rho}_h = \hat{\gamma}_h / \hat{\gamma}_0$ against the lag h. $\hat{\rho}_h$ takes values between -1 and 1 ($\hat{\rho}_h \in [-1, 1]$) and we always have $\hat{\rho}_0 = 1$.

Let $y_0, \ldots y_h$ be successive observations of a stationary time series and let $\tilde{y}_0$ and $\tilde{y}_h$ be such linear combinations of $y_1, \ldots, y_h$ that they minimize $E\{(y_0 - \tilde{y}_0)^2\}$ and $E\{(y_h - \tilde{y}_h)^2\}$, respectively. The partial autocorrelation function (PACF) is

$$\tilde{\rho}_1 = \mathrm{corr}(y_0, y_1), \qquad \tilde{\rho}_h = \mathrm{corr}(y_0 - \tilde{y}_0, y_h - \tilde{y}_h), \quad h = 2, \ldots. \tag{4}$$

The ACF shows the plain correlation between the data and its past value h time steps ago. The PACF, on the other hand, shows the conditional correlation between the data and its past value at lag h, given that the datapoints in the middle (at lags $1, 2, \ldots, h-1$) are held constant.

When illustrating the ACF and PACF we usually add dashed horizontal lines at $\pm 2/\sqrt{n}$, where n is the number of observations. If the ACF or PACF value at some lag h is outside these error boundaries this suggests that there is serial correlation at that lag, on the 5% significance level. However, 5% of the values are expected to be outside the boundaries even though the series really had no serial correlation.

A.3 ARMA models

Autoregressive moving average (ARMA) processes [Yule, 1926,Slutsky, 1927] are the most common set of time series models. They explain the behavior of the series by correlations to the previous values and previous errors in the same series. A time series $\{Y_t\}$ is an ARMA(p, q) process if

$$Y_t = \sum_{i=1}^{p} \phi_i Y_{t-i} + \varepsilon_t + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}, \tag{5}$$

where $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$ is a white noise series (see Appendix A.1) and ϕ_i and θ_i are constants. The first two terms of (5) form the autoregressive part of the model and the latter two terms the moving average part. An ARMA(p, q) model is weakly stationary if the parameters satisfy $|\phi_i| < 1$ for all $i \in \{1, \ldots, p\}$ and the roots of

$$\phi(L) = 0 \tag{6}$$

lie outside the unit circle. Here $\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p$ is the autoregressive operator. Thus, the stationarity of an ARMA process depends only on the parameters of the autoregressive part.

An ARMA(p, q) process can also be written with the lag operator ($Lx_t = x_{t-1}$) as follows

$$\phi(L) Y_t = \theta(L) \varepsilon_t, \tag{7}$$

where $\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p$ is the autoregressive operator and $\theta(L) = 1 + \theta_1 L + \theta_2 L^2 + \ldots + \theta_p L^p$ is the moving average operator.

A.4 Ljung-Box test

The Ljung-Box test [Ljung and Box, 1978] belongs to the family of Portmanteau tests and is used to determine whether a time series is just white noise. The test statistic in Ljung-Box test looks at the residual correlation at individual lags and takes into account their magnitudes as a group. The formulation is

$$Q(K) = n(n+2) \left(\frac{\hat{\rho}_1^2}{n-1} + \frac{\hat{\rho}_2^2}{n-2} + \ldots + \frac{\hat{\rho}_K^2}{n-K} \right), \tag{8}$$

where $\hat{\rho}_h$ is the empirical autocorrelation (see Appendix A.2) of the series at lag h, n is the sample size and K is the maximum number of lags being tested.

The null hypothesis is that the series is white noise, i.e., H_0: $\hat{\rho}_1 = \ldots \hat{\rho}_K = 0$, and the alternative hypothesis is H_1: $\hat{\rho}_i \neq 0$ for some $i \in \{1, \ldots, K\}$. Under the null hypothesis the test statistic Q has an approximate chi-square distribution with K degrees of freedom ($Q(K) \sim \chi^2$).

The decision rule is to reject H_0 if $Q(K) > \chi^2_\alpha$, where χ^2_α is the $100(1-\alpha)th$ percentile of the chi-square distribution with K degrees of freedom. The p-value, on the given significance level α, is then the probability of obtaining a test statistic at least as extreme as the one that was actually observed, assuming that the null hypothesis is true. Thus, if the Ljung-Box test provides a p-value smaller than the chosen significance level we have evidence against the series being uncorrelated for some lag $h < K$, at that significance level.

A.5 Information criteria

The Akaike information criterion is a way to compare the goodness of fit of different time series models [Akaike, 1973]. It is calculated

$$AIC = -2l(\hat{\boldsymbol{\theta}}; \boldsymbol{y}) + 2k, \tag{9}$$

where $l(\hat{\boldsymbol{\theta}}; \boldsymbol{y})$ is the log-likelihood of the model, $\hat{\boldsymbol{\theta}}$ is the maximum likelihood estimator and k is the number of parameters in the model. A larger likelihood means that the model is more likely to have produced the data than an other model with a lower likelihood. Adding parameters to a model always improves the likelihood, and thus we need a penalty term ($2k$ in (9)) to prevent too complicated models. The lower is the AIC the better fit the model is to the data.

The Bayesian information criterion is similar to the AIC, but it puts more weight on the penalty term especially with large datasets [Akaike, 1980]. The formula is

$$BIC = -2l(\hat{\boldsymbol{\theta}}; \boldsymbol{y}) + k\log(n), \tag{10}$$

where $l(\hat{\boldsymbol{\theta}}; \boldsymbol{y})$ is the log-likelihood with the maximum likelihood estimator $\hat{\boldsymbol{\theta}}$, k is the number of parameters and n the number of observations in the dataset.

A.6 Heteroskedasticity

Sometimes the error terms of a time series have different variances at different times even though they are uncorrelated with each other. In this case the series is said to have heteroskedastic effects [Hamilton, 1994]. On the other hand, if the variance of the error terms stays unchanged the series is called homoskedastic.

A.7 Models of heteroskedasticity

When the variance clearly does not stay the same at different parts of the series, there are probably heteroskedastic effects, i.e., the variance is time dependent and correlates with the previous variances. This kind of series can be modelled with autoregressive conditional heteroskedasticity (ARCH) models, generalized autoregressive conditional heteroskedasticity (GARCH) models, and many variations of these.

The ARMA processes (Appendix A.3) model the serial correlations in a time series whereas the processes of heteroskedasticity model the correlation of the variances. It is important to note the difference between unconditional and conditional variances. The unconditional variance of the residuals is a constant also for models of heteroskedasticity, but the conditional variance can change over time.

A.7.1 ARCH

The autoregressive conditional heteroskedasticity (ARCH) model introduced in [Engle, 1982] is the first and simplest of the heteroskedastic models. ARCH is a nonlinear, stationary time series model which allows the variance σ_t^2 to depend on the past values of the series. This makes it possible for the model to follow the natural pattern of volatility clustering. When the past variances are high the model estimates the next variance to be higher than it would be in case of low past variances. The coefficients in the ARCH model determine how fast the effect of past variances dies out.

A time series process $\{u_t\}$ is an ARCH(m) if the square of u_t can be described as an AR(m) process [Hamilton, 1994]:

$$u_t^2 = \zeta + \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2 + \ldots + \alpha_m u_{t-m}^2 + w_t, \tag{11}$$

where w_t is white noise with zero mean and

$$Var(w_t) = \lambda^2. \tag{12}$$

Another representation for an ARCH(m) process is

$$u_t = \sqrt{h_t} \cdot v_t, \tag{13}$$

where v_t is a white noise series with zero mean and unit variance. The coefficient h_t is calculated according to

$$h_t = \zeta + \sum_{j=1}^{m} \alpha_j u_{t-j}^2 = \zeta + A(L)u_t^2, \tag{14}$$

where $Lx_t = x_{t-1}$ is the lag operator and $A(L) = \sum_{j=1}^{m} \alpha_j L^j$ the coefficient operator. Both of these representations ((11) and (13)) lead to the conditional variance of u_t to be

$$E(u_t^2 \mid u_{t-1}^2, u_{t-2}^2, \ldots) = \zeta + \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2 + \ldots + \alpha_m u_{t-m}^2, \tag{15}$$

which depends on the past variances and can change over time. The unconditional variance, however, remains constant.

Since h_t must always be positive (because of the square root) the series w_t in (11) has to be bounded from below by $-\zeta$ and the coefficients must satisfy the conditions $\zeta > 0$ and $\alpha_j \geq 0,\ j = 1, \ldots, m$. The first coefficient ζ represents the long-term mean of the variance and the other coefficients α_j determine how fast the effects of a shock die out.

Stationarity is an important property of a time series process, since many time series models are derived for stationary series only. A process is covariance stationary if its mean and variance do not change over time and the correlation between two points depends only on the time distance between the points. An ARCH(m) process is covariance stationary (or weakly stationary) if the roots of

$$1 - \alpha_1 L - \alpha_2 L^2 - \ldots - \alpha_m L^m = 0 \tag{16}$$

lie outside the unit circle [Hamilton, 1994].

A.7.2 GARCH

The ARCH model was found useful but to get good fits a large number of parameters m was required. To overcome this problem another model type was developed. An ARCH(m) model allows the conditional variance to depend on the past sample variances but in a generalized autoregressive conditional heteroskedasticity GARCH(r, m) model the variances depend also on the lagged conditional variances [Bollerslev, 1986], thus less parameters are needed.

A time series process $\{u_t\}$ is a GARCH(r, m) if

$$u_t = \sqrt{h_t} \cdot v_t, \tag{17}$$

where v_t is a white noise series with zero mean and unit variance. The coefficient h_t is

$$\begin{aligned} h_t &= \kappa + \sum_{j=1}^{r} \beta_j h_{t-j} + \sum_{j=1}^{m} \alpha_j u_{t-j}^2 \\ &= \kappa + B(L)h_t + A(L)u_t^2, \end{aligned} \tag{18}$$

where $\kappa = (1 - \beta_1 - \beta_2 - \cdots - \beta_r)\zeta$. The latter form in (18) uses the lag operator L, i.e., $L_1 x_t = x_{t-1}$ and the coefficient operators $A(L) = \sum_{j=1}^{m} \alpha_j L^j$ and $B(L) = \sum_{j=1}^{r} \beta_j L^j$. It is required that $\zeta > 0$, $\alpha_j \geq 0, j = 1, \ldots, m$ and $\beta_j \geq 0, j = 1, \ldots, r$. These restrictions for the coefficients are needed to ensure that h_t is always positive.

A GARCH(r, m) process is covariance stationary if the roots of

$$1 - (\beta_1 + \alpha_1)L - (\beta_2 + \alpha_2)L^2 - \ldots - (\beta_p + \alpha_p)L^p = 0, \tag{19}$$

where p is the larger of r and m, are outside the unit circle [Hamilton, 1994].

Actually, a GARCH process relates to an ARCH process quite directly as GARCH can be shown to be equivalent to ARCH(m), where $m \to \infty$ [Hamilton, 1994]. Thus, a GARCH process is a generalization of an ARCH process.

A generalized ARCH process with infinite number of parameters is

$$h_t = \zeta + \pi(L)u_t^2, \tag{20}$$

where

$$\pi(L) = \sum_{j=1}^{\infty} \pi_j L^j. \tag{21}$$

This can be parameterized as the ratio of two finite-order polynomials [Hamilton, 1994]

$$\pi(L) = \frac{\alpha(L)}{1-\beta(L)} = \frac{\alpha_1 L + \alpha_2 L^2 + \cdots + \alpha_m L^m}{1 - \beta_1 L - \beta_2 L^2 - \cdots - \beta_r L^r}. \tag{22}$$

We assume that the roots of $1-\beta(L)=0$ lie outside the unit circle and thus we can multiply both sides of equation (20) by $1-\beta(L)$. The result is

$$[1-\beta(L)]h_t = [1-\beta(L)]\zeta + \alpha(L)u_t^2, \tag{23}$$

which is equivalent to

$$\begin{aligned} h_t &= \kappa + \beta_1 h_{t-1} + \beta_2 h_{t-2} + \cdots + \beta_r h_{t-r} \\ &\quad + \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2 + \cdots + \alpha_m u_{t-m}^2, \end{aligned} \tag{24}$$

where $\kappa = (1 - \beta_1 - \beta 2 - \cdots - \beta_r)\zeta$. This is exactly the definition of a GARCH(r,m) process presented earlier in (18).

GARCH connects also to ARMA models as the variance in a GARCH process follows an ARMA model. The following representation of a GARCH(r,m) model shows this connection.

$$u_t^2 = \kappa + \sum_{i=1}^{\max(m,r)} (\alpha_i + \beta_i)u_{t-i}^2 + \eta_t - \sum_{j=1}^{r} \beta_j \eta_{t-j}, \tag{25}$$

where $\eta = u_t^2 - h_t$. GARCH models have been proven quite effective and because of the possibility to depend also on past estimated variances, GARCH(1,1) model is often enough to model the heteroskedasticity in a time series.

A.7.3 Other models of heteroskedasticity

Even though ARCH and GARCH models are quite efficient they have some limitations which have inspired several more complicated models. First of all, ARCH and GARCH models require the coefficients to be non-negative, but in reality there can be negative correlations between current values and future volatility [Black, 1976]. This problem is solved by exponential GARCH (EGARCH) model [Nelson, 1991], where taking the logarithm ensures that the variance never becomes negative even with some negative coefficients. The formula for the h_t in (17) in an EGARCH process is [Hamilton, 1994]

$$\log h_t = \zeta + \sum_{j=1}^{\infty} \pi_j \cdot \{|v_{t-j}| - E|v_{t-j}| + \aleph v_{t-j}\}, \tag{26}$$

where v_t is normally distributed with zero mean and unit variance. This model is sometimes easier to estimate, compared to a GARCH model, since the h_t is always positive with any chosen parameters.

Another problem with ARCH and GARCH models is the symmetric reactions to positive and negative shocks, unlike what might happen in reality. The parameter $\aleph$ in (26) allows the model to react differently to these shocks. The issue is also solved with, e.g., the threshold GARCH (TGARCH) model [Zakoian, 1994], where an extra term is added to equation (18) to allow different behavior in the case of positive versus negative shock. The formulation of (18) for a TGARCH model is [Tsay, 2010]

$$h_t = \kappa + \sum_{i=1}^{r} (\alpha_i + \gamma_i N_{t-i}) u_{t-i}^2 + \sum_{j=1}^{m} \beta_j h_{t-j}, \tag{27}$$

where κ, α_i, β_i and γ_i are nonnegative parameters satisfying similar conditions as those of GARCH models. N_{t-i} is an indicator for negative u_{t-i}

$$N_{t-i} = \begin{cases} 1 & \text{if } u_{t-i} < 0 \\ 0 & \text{if } u_{t-i} \geq 0 \end{cases} . \tag{28}$$

From among the numerous GARCH model types we mention also the GARCH-in-mean (GARCH-M) model [Engle et al., 1987] where a heteroskedasticity term is added also to the main equation (17), and the integrated GARCH (IGARCH) model where the coefficients in (18) are restricted to satisfy [Bollerslev, 1986]

$$\sum_{j=1}^{m} \alpha_j + \sum_{j=1}^{r} \beta_j = 1. \tag{29}$$

This limitation ensures that the GARCH process has a unit root.

A.8 Estimation method for ARCH and GARCH models

The simple heteroskedastic ARCH and GARCH models are usually estimated using the maximum likelihood approach [Bera and Higgins, 1993].

If we assume that the white noise series v_t in (13) is Normally distributed then the conditional likelihood function of an ARCH(m) model is

$$f(u_{m+1}, \ldots, u_T|\boldsymbol{\alpha}, u_1, \ldots, u_m) = \prod_{t=m+1}^{T} \frac{1}{\sqrt{2\pi h_t}} \exp\left(-\frac{u_t^2}{2h_t}\right), \tag{30}$$

where $\boldsymbol{\alpha} = (\zeta, \alpha_1, \ldots, \alpha_m)'$ is a vector consisting of the coefficients in the model [Tsay, 2010].

Maximizing the likelihood function is equivalent to maximizing its logarithm, which is usually easier to compute. The conditional log-likelihood function of an ARCH(m) model is

$$l(u_{m+1}, \ldots, u_T|\boldsymbol{\alpha}, u_1, \ldots, u_m) = \sum_{t=m+1}^{T} \left[-\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(h_t) - \frac{1}{2}\frac{u_t^2}{h_t}\right]. \tag{31}$$

The first term in (31) is a constant and the log-likelihood function becomes

$$l(u_{m+1}, \ldots, u_T|\boldsymbol{\alpha}, u_1, \ldots, u_m) = -\sum_{t=m+1}^{T} \left[\frac{1}{2}\log(h_t) + \frac{1}{2}\frac{u_t^2}{h_t}\right]. \tag{32}$$

Then again, if the white noise series v_t is assumed to be t-distributed with w degrees of freedom, then the conditional log-likelihood function takes the form [Tsay, 2010]

$$l(u_{m+1}, \ldots, u_T|\boldsymbol{\alpha}, U_m) = -\sum_{t=m+1}^{T} \left[\frac{w+1}{2}\log\left(1 + \frac{u_t^2}{(w-2)h_t}\right) + \frac{1}{2}\log(h_t)\right], \tag{33}$$

where $U_m = u_1, \ldots, u_m$.

Once the log-likelihood function has been determined the correct parameter values are found by calculating the maximum likelihood estimate. It is numerically computed by a Quasi-Newton optimizer.

To estimate GARCH models a two-pass estimation method can be used [Tsay, 2010]. This means that the parameters are estimated using the ARMA representation of the squared series in formula (25) and the maximum likelihood method. Let the AR and MA coefficient estimates be $\hat{\phi}_i$ and $\hat{\theta}_i$. The GARCH estimates are then $\hat{\beta}_i = \hat{\theta}_i$ and $\hat{\alpha}_i = \hat{\phi}_i - \hat{\theta}_i$.

A.9 LM-test for heteroskedasticity

The Lagrange multiplier test is an often used tool for detecting heteroskedasticity in a time series. The squared values of the series $\{u_t\}$ are regressed on a constant and m of its own lagged values:

$$u_t^2 = \zeta + \alpha_1 u_{t-1}^2 + \alpha_2 u_{t-2}^2 + \ldots + \alpha_m u_{t-m}^2 + e_t, \tag{34}$$

for $t = 1, 2, \ldots, T$, where T is the sample size [Hamilton, 1994].

The null hypothesis is that there are no heteroskedastic effects (H_0: $\alpha_1 = \alpha_2 = \ldots = \alpha_m = 0$), i.e., no volatility clustering or correlation in the variances. The alternative hypothesis is that there exists i s.t. $\alpha_i \neq 0$.

The test statistic is the usual F-statistic for the regression on the squared series which, under the null hypothesis, follows a χ^2-distribution with m degrees of freedom. Thus, if the test provides a p-value smaller than the chosen significance level, there is evidence against the null hypothesis and there probably is heteroskedasticity in the data.

References

[Akaike, 1973] Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In Kotz, S. and Johnson, N. L., editors, *Breakthroughs in Statistics*, volume 1, pages 610–624. Springer, New York.

[Akaike, 1980] Akaike, H. (1980). Seasonal adjustment by a bayesian modeling. *Journal of Time Series Analysis*, 1(1):1–13.

[Ball and Mankiw, 1994] Ball, L. and Mankiw, N. G. (1994). Asymmetric price adjustment and economic fluctuations. *The Economic Journal*, 104(423):247–261.

[Bekaert and Wu, 2000] Bekaert, G. and Wu, G. (2000). Asymmetric volatility and risk in equity markets. *Review of Financial Studies*, 13(1):1–42.

[Bera and Higgins, 1993] Bera, A. K. and Higgins, M. L. (1993). Arch models: Properties, estimation and testing. *Journal of economic surveys*, 7(4):305–366.

[Black, 1976] Black, F. (1976). Studies of stock market volatility changes. *Proceedings of the American Statistical Association, Business and Economic Statistics Section*, pages 177–181.

[Bollerslev, 1986] Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31:307–327.

[Cover, 1992] Cover, J. P. (1992). Asymmetric effects of positive and negative money-supply shocks. *The Quarterly Journal of Economics*, 107(4):1261–1282.

[Davison, 2011] Davison, A. C. (2011). *Time Series*. Ecole Polytechnique Federale de Lausanne, Lausanne, new edition.

[Engle, 1982] Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 50(4):987–1007.

[Engle and Kroner, 1995] Engle, R. F. and Kroner, K. F. (1995). Multivariate simultaneous generalized arch. *Econometric Theory*, 11(1):122–150.

[Engle et al., 1987] Engle, R. F., Lilien, D. M., and Robins, R. P. (1987). Time varying risk premia in the term structure The arch-m model. *Econometrica*, 55(2):391–407.

[Hamilton, 1994] Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press, Chichester, West Sussex.

[Hendry and Juselius, 2000] Hendry, D. F. and Juselius, K. (2000). Explaining cointegration analysis: Part 1. *The Energy Journal*, 21(1):1–42.

[Ljung and Box, 1978] Ljung, G. M. and Box, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrica*, 65(2):297–303.

[Loretan and Phillips, 1994] Loretan, M. and Phillips, P. C. B. (1994). Testing the covariance stationarity of heavy-tailed time series: An overview of the theory with applications to several financial datasets. *Journal of Empirical Finance*, 1(2):211–248.

[Nelson, 1991] Nelson, D. B. (1991). Heteroskedasticity in asset returns: A new approach. *Econometrica*, 59(2):347–370.

[Parkin et al., 2008] Parkin, M., Powell, M., and Matthews, K. (2008). *Economics*. Pearson Education Limited, Essex, seventh edition.

[Searls, 1964] Searls, D. T. (1964). The utilization of a known coefficient of variation in the estimation procedure. *Journal of the American Statistical Association*, 59(308):1225–1226.

[Slutsky, 1927] Slutsky, E. (1927). The summation of random causes as the source of cyclic processes. *Econometrica*, 5:105–146.

[Tsay, 2010] Tsay, R. S. (2010). *Analysis of Financial Time Series*. John Wiley & Sons, Inc, Hoboken, New Jersey, third edition.

[Yule, 1926] Yule, G. U. (1926). Why do we sometimes get nonsense-correlations between time-series? - a study in sampling and the nature of time-series. *Journal of the Royal Statistical Society*, 89:1–63.

[Zakoian, 1994] Zakoian, J.-M. (1994). Threshold heteroskedastic models. *Journal of Economic Dynamics and Control*, 18(5):931–955.